



REVIEW

The changing epitome of species identification – DNA barcoding



M. Ajmal Ali ^{a,*}, Gábor Gyulai ^b, Norbert Hidvégi ^b, Balázs Kerti ^b,
Fahad M.A. Al Hemaïd ^a, Arun K. Pandey ^c, Joongku Lee ^d

^a Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

^b Institute of Genetics and Biotechnology, St. István University, Gödöllo H-2103, Hungary

^c Department of Botany, University of Delhi, Delhi 110007, India

^d International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, 111 Gwahangno, Yuseong-gu, Daejeon 305 806, South Korea

Received 17 February 2014; revised 22 March 2014; accepted 23 March 2014

Available online 31 March 2014

KEYWORDS

DNA barcoding;
Molecular markers;
Species identification;
Plant taxonomy;
Biodiversity;
Conservation genetics

Abstract The discipline taxonomy (the science of naming and classifying organisms, the original bioinformatics and a basis for all biology) is fundamentally important in ensuring the quality of life of future human generation on the earth; yet over the past few decades, the teaching and research funding in taxonomy have declined because of its classical way of practice which lead the discipline many a times to a subject of opinion, and this ultimately gave birth to several problems and challenges, and therefore the taxonomist became an endangered race in the era of genomics. Now taxonomy suddenly became fashionable again due to revolutionary approaches in taxonomy called DNA barcoding (a novel technology to provide rapid, accurate, and automated species identifications using short orthologous DNA sequences). In DNA barcoding, complete data set can be obtained from a single specimen irrespective to morphological or life stage characters. The core idea of DNA barcoding is based on the fact that the highly conserved stretches of DNA, either coding or non coding regions, vary at very minor degree during the evolution within the species. Sequences suggested to be useful in DNA barcoding include cytoplasmic mitochondrial DNA (e.g. *cox1*) and chloroplast DNA (e.g. *rbcL*, *trnL-F*, *matK*, *ndhF*, and *atpB rbcL*), and nuclear DNA (ITS, and house keeping genes e.g. *gapdh*). The plant DNA barcoding is now transitioning the epitome of species identification; and thus, ultimately helping in the molecularization of taxonomy, a need

* Corresponding author. Tel.: +966 11 46 75734.

E-mail addresses: alimohammad@ksu.edu.sa (M. Ajmal Ali), Gyulai.Gabor@mkk.szie.hu (G. Gyulai), Hidvegi.Norbert@mkk.szie.hu (N. Hidvégi), Kerti.Balasz@mkk.szie.hu (B. Kerti), fhemaïd@ksu.edu.sa (F.M.A. Al Hemaïd), arunpandey79@gmail.com (A.K. Pandey), joongku@kribb.re.kr (J. Lee).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

of the hour. The ‘DNA barcodes’ show promise in providing a practical, standardized, species-level identification tool that can be used for biodiversity assessment, life history and ecological studies, forensic analysis, and many more.

© 2014 Production and hosting by Elsevier B.V. on behalf of King Saud University.

Contents

1. Introduction	205
2. Molecular phylogeny and DNA barcoding	207
2.1. Ribosomal DNA (rDNA) of the nuclear genome (nuDNA) – ITS	208
2.2. Chloroplast DNA (cpDNA)	211
2.3. Mitochondrial DNA (mtDNA)	211
3. Molecular phylogenetic analyses	213
4. Retrotransposon based barcoding	215
5. Metabarcoding	219
6. Materials for DNA barcoding	219
7. Benefits of DNA barcoding	219
7.1. Plant systematics	220
7.2. Medicinal and wild plants’ identification	220
7.3. Food safety and conservation biology	221
8. Limitations of DNA barcoding	221
9. Conclusions	222
Acknowledgements	222
References	222

1. Introduction

There are approximately 1.7 million species identified by using morphological (*i.e.* Linnean) characters including 808 gymnosperm, and 90,000 monocots and about 200,000 dicots of angiosperm. This number may be a gross under-estimate of the true biological diversity of Earth (Blaxter, 2003; Wilson, 2003). Recently, overwhelming landmark publications (Table 1) on DNA barcoding (Hebert et al., 2003) (*syn.*: profiling, genotyping) based on highly conserved sequence informations provide new tools for systematics (Hebert and Barrett, 2005) and phylogeny (Wyman et al., 2004; Leebens-Mack et al., 2005; Jansen et al., 2006; Hansen et al., 2006). DNA barcodes consist of short sequences of DNA between 400 and 800 base pairs that can be routinely amplified by PCR (polymerase chain reaction) and sequenced of the species studied.

Morphologically distinguishable taxa may not require barcoding; however, subspecies (*ssp.*), cultivars (*cv.*), eco- and morphotypes, mutants, species complex and clones can be diagnosed with molecular barcoding. Barcode of a specimen can be compared with sequences derived from other taxa, and in the case of dissimilarities species identity can be determined by molecular phylogenetic analyses based on MOTU, molecular operational taxonomic units (Floyd et al., 2002).

DNA barcoding was particularly useful for marine organisms (Shander and Willassen, 2005), including fishes (Mason, 2003; Ward et al., 2005); soil meiofauna (Blaxter et al., 2004) and freshwater meiobenthos (Markmann and Tautz, 2005); and extinct birds (Lambert et al., 2005). In the rainforests, rapid DNA-based entomological inventories were so effective (Monaghan et al., 2005; Smith et al., 2005) that tropical ecologists were the most active advocates of DNA barcoding

(Janzen, 2004). More pragmatically, DNA barcodes have proved to be useful in biosecurity, *e.g.* for surveillance of disease vectors (Besansky et al., 2003) and invasive insects (Armstrong and Ball, 2005), as well as for law enforcement and primatology (Lorenz et al., 2005).

Barcoding has created some controversy in the taxonomy community (Mallet and Willmott, 2003; Lipscomb et al., 2003; Seberg et al., 2003; DeSalle et al., 2005; Lee, 2004; Ebach and Holdrege, 2005; Will et al., 2005). Traditional taxonomists use multiple morphological traits to delineate species. Today, such traits are increasingly being supplemented with DNA-based information. In contrast, the DNA barcoding identification system is based on what is in essence a single complex character (a portion of one gene, comprising ~650 bp from the first half of the mitochondrial cytochrome c oxidase subunit I gene sometimes called COXI or COI), and barcoding results are therefore seen as being unreliable and prone to errors in identification (Dasmahapatra and Mallet, 2006). Although the mitochondrial cytochrome oxidase subunit I (COI) is a widely used barcode in a range of animal groups (Hebert et al., 2003), this locus is unsuitable for use in plants due to its low mutation rate (Kress et al., 2005; Cowen et al., 2006; Fazekas et al., 2008). In addition, complex evolutionary processes, such as hybridization and polyploidy, are common in plants, making species boundaries difficult to define (Rieseberg et al., 2006; Fazekas et al., 2009).

The number and identity of DNA sequences that should be used for barcoding is a matter of debate (Pennisi, 2007; Ledford, 2008). The main DNA barcoding bodies and resources are (1) Consortium for the Barcode of Life (CBOL) <http://www.cbol.org>

Table 1 Some landmark articles related to DNA barcoding published onward 2003.

Year	Article	References
2003	Taxonomy, DNA, and the bar code of life Biological identifications through DNA barcodes A plea for DNA taxonomy	Stoeckle (2003) Hebert et al. (2003) Tautz et al. (2003)
2004	Now is the time DNA barcoding: promise and pitfalls Identification of birds through DNA barcodes Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification	Janzen (2004) Moritz and Cicero (2004) Hebert et al. (2004) Will and Rubinoff (2004)
2005	Genome sequencing in microfabricated high density picolitre reactors Toward writing the encyclopedia of life: an introduction to DNA barcoding DNA barcodes for Biosecurity: invasive species identification DNA barcoding does not compete with taxonomy Will DNA bar codes breathe life into classification? The promise of DNA barcoding for taxonomy DNA barcoding is no substitute for taxonomy The perils of DNA barcoding and the need for integrative taxonomy Emerging technologies in DNA sequencing Critical factors for assembling a high volume of DNA barcodes DNA barcoding: Error rates based on comprehensive sampling Nextgeneration DNA sequencing techniques	Margulies et al. (2005) Savolainen et al. (2005) Armstrong and Ball (2005) Gregory (2005) Marshall (2005) Hebert and Gregory (2005) Ebach and Holdrege (2005) Will et al. (2005) Metzker (2005) Hajibabaei et al. (2005) Meyer and Paulay (2005) Ansorge (2009)
2006	Who will actually use DNA barcoding and what will it cost? A minimalist barcode can identify a specimen whose DNA is degraded	Cameron et al. (2006) Hajibabaei et al. (2006)
2007	Limited performance of DNA barcoding in a diverse community of tropical butterflies A proposal for a standardized protocol to barcode all land plants	Elias et al. (2007) Chase et al. (2007)
2008	The impact of next-generation sequencing technology on genetics A universal DNA mini-barcode for biodiversity analysis DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics DNA barcodes: genes, genomics, and bioinformatics	Mardis (2008) Meusnier et al. (2008) Hajibabaei et al. (2008) Kress and Erickson (2008)
2009	Evaluation of next generation sequencing platforms for population targeted sequencing studies DNA barcoding for ecologists	Harismendy et al. (2009) Valentini et al. (2009)
2010	DNA barcoding: a six-question tour to improve users' awareness about the method A survey of sequence alignment algorithms for next-generation sequencing. Briefings in bioinformatics	Casiraghi et al. (2010) Li and Homer (2010)
2011	Pyrosequencing for mini-barcoding of fresh and old museum specimens Use of rbcL and trnL-F as a Two-Locus DNA Barcode for Identification of NWEuropean Ferns: An Ecological Perspective Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos On the future of genomic data	Shokralla et al. (2011) De Groot et al. (2011) Hajibabaei et al. (2011) Kahn (2011)
2012	An emergent science on the brink of irrelevance: a review of the past 9 years of DNA barcoding Next-generation sequencing technologies for environmental DNA research Environmental DNA Toward next-generation biodiversity assessment using DNA metabarcoding The golden age of metasytematics A bloody boon for conservation The future of environmental DNA in ecology Tracking earthworm communities from soil DNA Persistence of environmental DNA in freshwater ecosystems New environmental metabarcodes for analyzing soil DNA: potential for studying past and present ecosystems Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? Bioinformatic challenges for DNA metabarcoding of plants and animals ABGD, Automatic Barcode Gap Discovery for primary species delimitation Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies	Taylor and Harris (2012) Shokralla et al. (2012) Taberlet et al. (2012a) Taberlet et al. (2012c) Hajibabaei (2012) Callaway (2012) Yoccoz (2012) Bienert et al. (2012) Dejean et al. (2011) Epp et al. (2012) Carlsen et al. (2012) Coissac et al. (2012) Puillandre et al. (2012) Taberlet et al. (2012b)

Table 1 (continued)		
Year	Article	References
2013	DNA barcoding as a complementary tool for conservation and valorization of forest resources	Laiou et al. (2013)
	Incorporating trnH-psbA to the core DNA barcodes improves significantly species discrimination within southern African Combretaceae	Gere et al. (2013)
	A DNA mini-barcode for land plants. Mol Ecol Resour	Little (2013)
	Toward a unified paradigm for sequence-based identification of fungi	Kõljalg et al. (2013)
	Potential of DNA barcoding for detecting quarantine fungi. Phytopathology	Gao and Zhang (2013)
	DNA barcoding in plants: evolution and applications of in silico approaches and resources	Bhargava and Sharma (2013)
	The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS	Han et al. (2013)
	Assessing DNA barcoding as a tool for species identification and data quality control	Shen et al. (2013)
	The seven deadly sins of DNA barcoding	Collins and Cruickshank (2013)
	Use of the potential DNA barcode ITS2 to identify herbal materials	Pang et al. (2013)
2014	20 years since the introduction of DNA barcoding: from theory to application	Fišer and Buzan (2014)
	Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead	Joly et al. (2014)

barcodeoflife.org established in 2004. CBOL promotes DNA barcoding through over 200 member organizations from 50 countries, operates out of the Smithsonian Institution's National Museum of Natural History in Washington, (2) International Barcode of Life (iBOL) <http://www.ibol.org> launched in October 2010, iBOL represents a not-for-profit effort to involve both developing and developed countries in the global barcoding effort, establishing commitments and working groups in 25 countries. The Biodiversity Institute of Ontario is the project's scientific hub and its director, (3) The Barcode of Life Datasystems (BOLD) <http://www.boldsystems.org>. The Barcode of Life Datasystems is an online workbench for DNA barcoders, combines a barcode repository, analytical tools, interface for submission of sequences to GenBank, a species identification tool and connectivity for external web developers and bioinformaticians. It is established in 2005 by the Biodiversity Institute of Ontario. The Consortium for the Barcode of Life (CBOL) Plant Working Group (2009) recommended *rbcL* + *matK* as a core two-locus combination. However, as these loci encode conserved functional traits it is not clear whether they provide sufficiently high species resolution. One of the challenges for plant barcoding is the ability to distinguish closely related or recently evolved species. Recently plant DNA barcoding has focused on several studies (e.g. Lágler et al., 2006; Chase et al., 2007; Kress and Erickson, 2008; Edwards et al., 2008; Newmaster et al., 2008; Newmaster and Ragupathy, 2009; Seberg and Petersen, 2009; Spooner, 2009; Starr et al., 2009; Zhang et al., 2009; Mansour et al., 2009a,b; Clerc-Blain et al., 2010; Kelly et al., 2010; Zuo et al., 2011; Dong et al., 2011; Du et al., 2011; Fu et al., 2011; Gu et al., 2011; Guo et al., 2011; Li et al., 2011a,b; Shi et al., 2011; Wang et al., 2011; Xiang et al., 2011; Xue and Li, 2011; Yan et al., 2011; Yu et al., 2011; Liu et al., 2012; Saarela et al., 2013; Techen et al., 2014).

2. Molecular phylogeny and DNA barcoding

Dobzhansky (1973) stated that nothing in biology makes sense except in the light of evolution. Phylogeny is in the midst of a renaissance, heralded by the widespread application of new analytical approaches and molecular techniques. Phylogenetic

analyses provided insights into relationships at all levels of evolution. The phylogenetic trees now available at all levels of the taxonomic hierarchy for animals and plants, which play a pivotal role in comparative studies in diverse fields from ecology to molecular evolution and comparative genetics (Soltis and Soltis, 2000). The basic DNA nucleotide substitution rate was estimated to be 1.3×10^{-8} (Ma and Benetzen, 2004) and 6.5×10^{-9} (Gaut et al., 1996) substitution per locus per year in grasses, and it was estimated to 1.5×10^{-8} in *Arabidopsis* (Koch et al., 2000).

Traditionally, the studies of variation of a species are based on morphological characters. Variations at molecular level are primarily based on the changes of DNA nucleotide sequences of homologous genes of the populations of a species and higher taxa (Hamby and Zimmer, 1992). Our understanding on genetic variability of organisms located at various levels of the tree of life has advanced greatly with the advancement of the molecular techniques (Avisé, 1994; Hillis et al., 1996; Soltis et al., 1998; Hollingsworth et al., 1999; Wen and Pandey, 2005; Mondini et al., 2009). Plant genomes range in size of 8.8×10^6 to more than 300×10^9 bp, however DNA can be prepared from a small amount of leaf tissue (0.1 g). A diverse array of molecular techniques are available for studying genetic variability, including restriction site analysis, analysis of DNA rearrangements, gene and intron loss, and the dominantly used PCR based techniques followed by DNA sequencing and cladistic analyses of the nuclear genome (nuDNA) and both organelle genomes of mitochondria (mtDNA) and chloroplast (cpDNA) (Martins and Hellwig, 2005; Mitchell and Wen, 2005). Multiple sequence alignments software programs of BioEdit Sequence Alignment Editor (North Carolina State University, USA) (Hall, 1999), MULTALIN (Combet et al., 2000), CLUSTAL W (Thompson et al., 1997), FastPCR (Kalendar et al., 2009), BLAST analysis of the NCBI databases (Altschul et al., 1997) and MEGA5 (Tamura et al., 2011) are available for inferring Phylogeny.

The use of DNA or protein sequences to identify organisms was proposed as a more efficient approach than traditional taxonomic practices (Blaxter et al., 2004; Tautz et al., 2003). A chloroplast gene such as *matK* (maturase K) or a nuclear gene such as ITS (internal transcribed spacer) may be an

effective target for barcoding in plants (Kress et al., 2005). Kress et al. (2005) have demonstrated the effectiveness of DNA barcoding in angiosperms. Ribosomal DNA (e.g. ITS) could be used to complement of results based on plastid genes, that may provide a more sophisticated multiple component barcode for species diagnosis and delimitation (Chase et al., 2005). Sequences used for molecular barcoding are the nuclear small subunit ribosomal RNA gene (SSU, also known as 16S in prokaryotes, and 18S in most eukaryotes), the nuclear large-subunit ribosomal RNA gene (LSU, also known as 23S and 28S), the highly variable internal transcribed spacer section of the ribosomal RNA cistron (ITS, separated by the 5S ribosomal RNA gene into ITS1 and ITS2 regions), the mitochondrial cytochrome *c* oxidase 1 (CO1 or *cox1*) gene and the chloroplast ribulose biphosphate carboxylase large subunit (*rbcL*) gene. Kress et al. (2005) have suggested that ITS spacer region and the plastid *trnH-psbA* have greater potential for species-level discrimination than any other locus, the *trnH-psbA* combined with *rp136-rpf8*, and *trnL-F* ranked the highest amplification success with appropriate sequence length (Kress et al., 2005).

2.1. Ribosomal DNA (rDNA) of the nuclear genome (nuDNA) – ITS

Sequence analyses of the nuclear multicopy ribosomal DNA (rDNA) genes encoding for structural RNAs (rRNAs) of ribosomes have been widely used in plant phylogenetics (Baldwin, 1992; Baldwin et al., 1995; Hershkovitz et al., 1999). The rDNA is arranged in tandem repeats in one or a few chromosomal loci with thousands of repeats. In total, rDNA can comprise as much as 10% of the total plant genome (Zimmer et al., 1980). The size of repeating unit of the human rDNA (18S rRNA; 28S rRNA; 5.8S rRNA; 5'ETS; 3'ETS; ITS1; ITS2; intergenic spacer; *cdc27* pseudogene; p53 binding site) is 42.999 bp (NCBI # U13369).

The majority of angiosperm genome sizes have a narrow range between 135 and 160 kb (Fig. 1). In plant genomes,

the rDNA cistron encodes 18S, 26S and 5.8S rRNAs, which are separated by the two internal transcribed spacers (ITS1 and ITS2), and the cistron is flanked by the 5' and 3' external transcribed spacers (5'-ETS and 3'-ETS) (Fig. 2). The regions are relatively short sequences, ITS1 200–300 bp, ITS2 180–240 bp, and 5.8S 160 bp in flowering plants. The amplification and sequencing primers are highly universal (White et al., 1990). Eukaryotic ribosomes consist of two subunits with the large subunit (LSU), which is about twice the size of the small subunit (SSU). The 18S gene encodes the SSU; and LSU is encoded by 26S and 5.8S. The ITS region comprises the 5.8S gene, which has been the most widely used molecular marker at interspecific and intergeneric levels (Feliner et al., 2004). Because of the influence of concerted evolution (Zimmer et al., 1980), the polymorphisms are not due to the intra-genomic variability at these loci, rather, a more frequent merge of different ITS copies within the same genome (Campbell et al., 1997; Buckler et al., 1997; Hershkovitz et al., 1999; Feliner et al., 2004).

The 18S gene is a slowly evolving marker and is also suitable for inferring phylogenies of angiosperms (Hamby and Zimmer, 1992; Soltis et al., 1997), and closely related families such as Caryophyllales (Cuenoud et al., 2002); however, the most common limitation of 18S rDNA using for phylogenetic analyses, is its low levels of variability within the angiosperms.

The phylogenetic utility of the 26S sequences has not been widely explored. In plants, the 26S gene is about 3.4 kb long and includes 12 expansion segments (ES), which are variable (Bult et al., 1995). The overall nucleotide substitution rate of 26S is 1.6–2.2 times higher than that in 18S (Kuzoff et al., 1998). The 26S sequences have been used to discriminate closely related families such as of Apiales (Chandler and Plunkett, 2004), and for determining phylogenetic position of plant families (Simmons et al., 2001; Neyland, 2002).

The external transcribed spacer (ETS) region (especially the 3' end of the 5'-ETS sequence adjacent to 18S) has also been used for phylogenetic analyses. The frequency of ETS polymorphism is still poorly understood. Furthermore, the

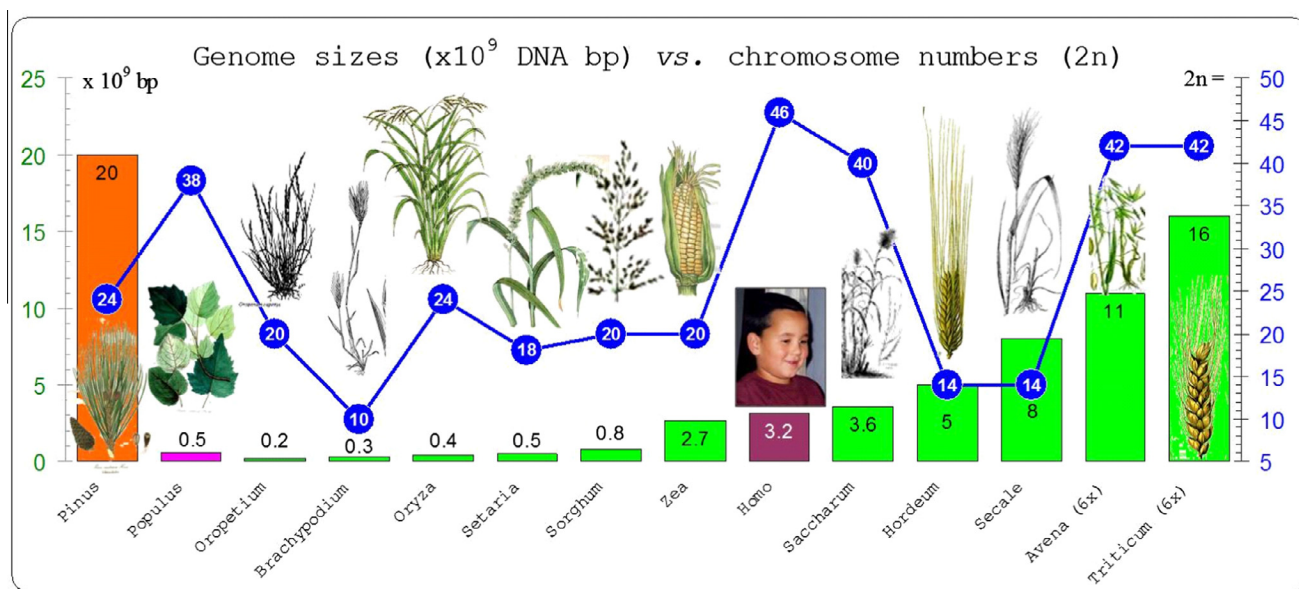


Figure 1 Sizes of plant genomes. Sizes (bp) vs. chromosome numbers (2n) of plant genomes from different taxa and compared to human genome size (3.2×10^9 bp).

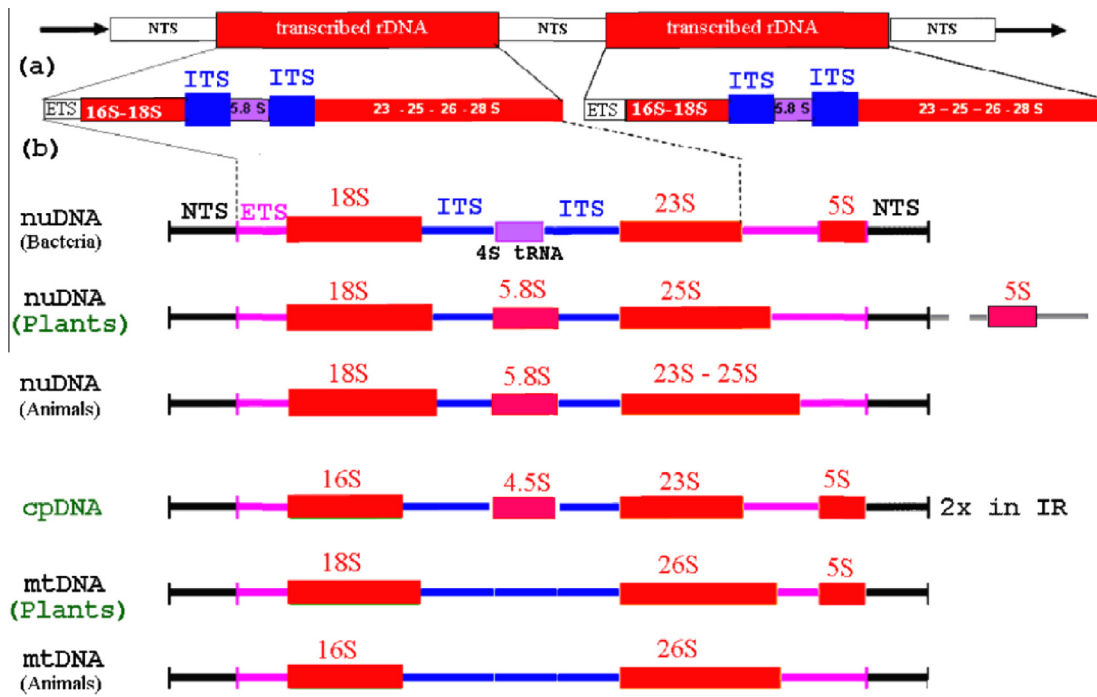


Figure 2 rDNA sequence domains of tandem gene clusters (about 10 kb each) of nuclear and organellar rDNAs of different organisms and organelles (cp and mt). (a) Pre-rDNA in the genome and (b) the translated ribosomal subunits. *Abbreviations:* ETS – external transcribed spacer, ITS – internal transcribed spacer, NTS – non transcribed spacer, mt – mitochondrion; cp – chloroplast, nu – nucleus, 5.8S to 28S – ribosomal subunits of nucleoproteins of rRNAs.

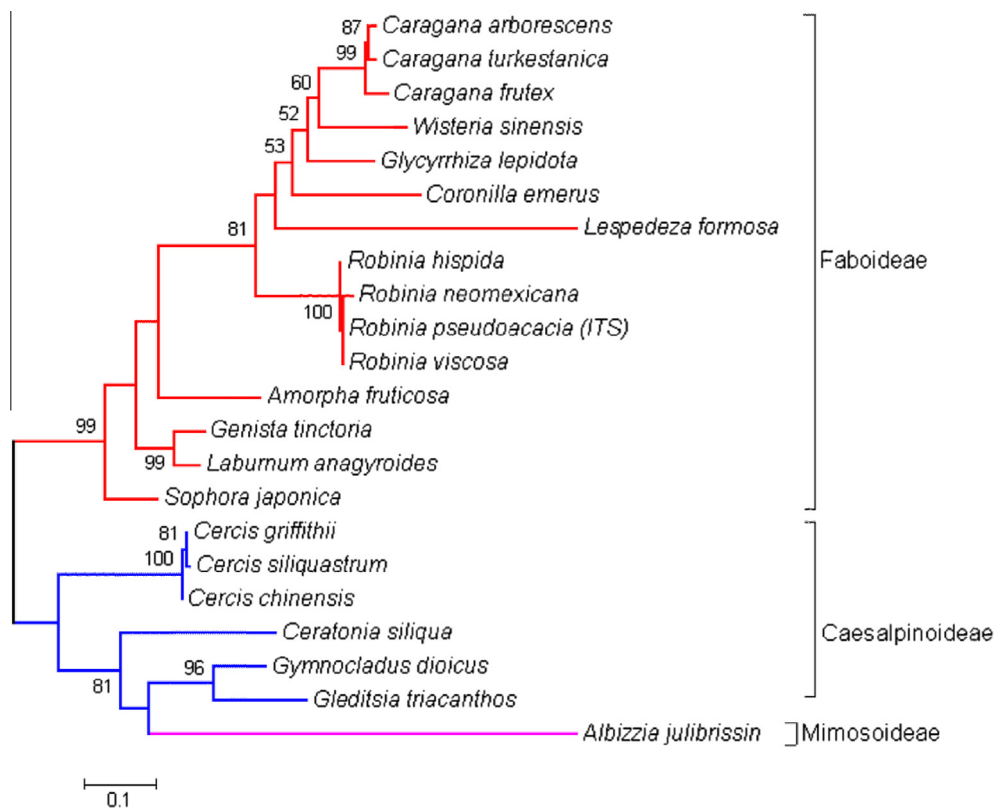


Figure 3 Sample of ITS phylogeny. ITS cladogram of Legume trees. ITS sequences of twenty-two species were analyzed by BioEdit (Hall, 1999), and ML (Maximum Likelihood; Hillis et al., 1994) cladogram was edited by MEGA4 (Tamura et al., 2007) (×1000 bootstrap). The three subfamilies of Fabaceae and the substitution rate (0.1) are indicated.

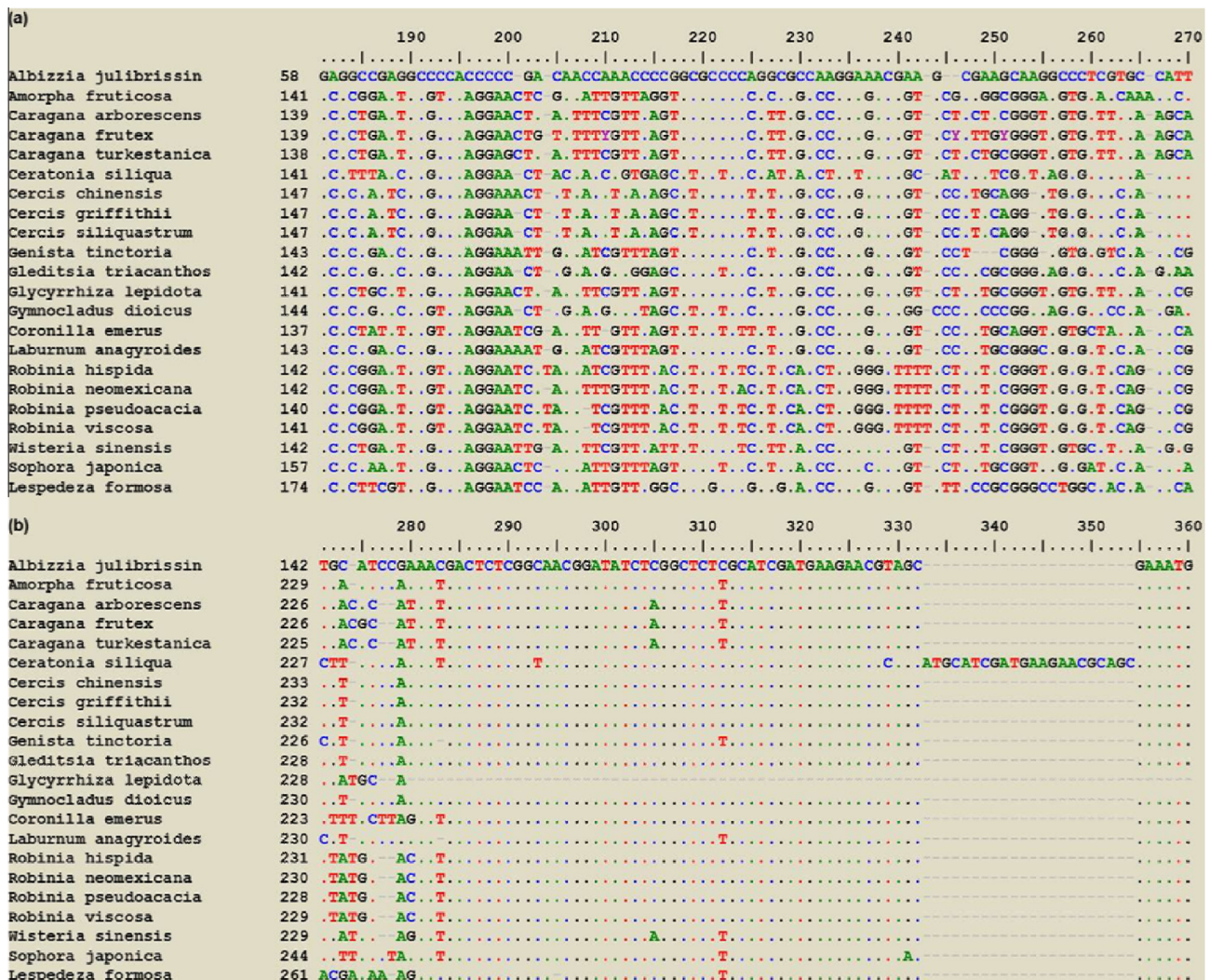


Figure 4 Samples of ITS sequence polymorphism. Sequence alignment of ITS sequences of rDNAs of twenty-two Legume trees studied with low (a) (180–270 nt) and high (b) (270–360 nt) sequence similarities.

internal repetitive structure of the ETS region can make the amplifications and sequence alignments difficult (Baldwin and Markos, 1998; Linder et al., 2000).

To compare all rDNA loci, the ITS is the most commonly sequenced locus used in plant phylogenetic studies (Pandey and Ali, 2006, 2012; Pandey et al., 2009; Choudhary et al., 2011; Ali et al., 2012, 2013; Lee et al., 2013) (Figs. 3 and 4). The advantage of the ITS region is that it can be amplified in two smaller fragments (ITS1 and ITS2) with the joining 5.8S locus. The quite conserved 5.8S region, in fact, contains enough phylogenetic signals for discrimination at levels of orders and phyla, however this locus is not the concern of barcoding. ITS regions often vary by insertions and deletions within an individual rather than substitution, which makes sequencing difficult as several ITS sequence types are being analyzed simultaneously (Elbadri et al., 2002).

Taxonomy is a synthetic science, drawing upon data from such diverse fields as morphology, anatomy, embryology, cytology, and chemistry. In recent years, development of techniques in molecular biology including those for molecular hybridization, cloning, restriction endonuclease digestions and nucleic acid sequencing have provided many new tools

for the investigation of phylogenetic relationships. The reconstructions of angiosperm phylogeny have relied largely on plastid and mitochondrial genes (Chase et al., 1993; Nandi et al., 1998; Savolainen et al., 2000; Hilu et al., 2003; Zhu et al., 2007; Qiu et al., 2010) and sometimes entire plastid genomes (Jansen et al., 2007; Moore et al., 2007, 2010) and nuclear genes (Doyle et al., 1994; Soltis et al., 1997; Mathews and Donoghue, 1999; Finet et al., 2010; Lee et al., 2011). Moreover, our understanding of the relationships among organisms at various levels in the tree of life has been advanced greatly in the last about three decades with the aid of DNA molecular systematic techniques and phylogenetic theory, this has resulted into a classification ‘Angiosperm Phylogeny Group’ (APGI, 1998; APGII, 2003; APGIII, 2009; Haston et al., 2009) of the families of the flowering plants (<http://www.mobot.org/MOBOT/research/APweb/>). From the first report of the utility of the internal transcribed spacers (ITS) sequence of nuclear ribosomal DNA (nrDNA) in plants (Baldwin, 1992), it has been extensively used to distinguish even very closely related species (Chen et al., 2010; Yao et al., 2010). Moreover, in the last two decades, the ITS sequence has gained much attention, along with the smartest genes available for

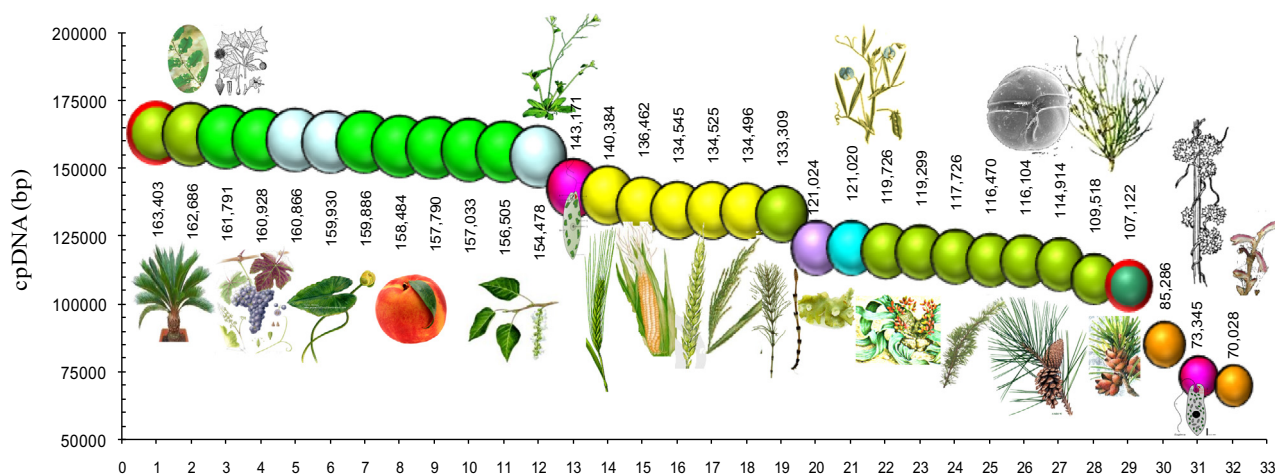


Figure 5 cpDNA. Comparative sizes (bp) of total plant cpDNAs of different taxa (with NCBI accession numbers; Altschul et al., 1997). (1) Gymnosperm *Cycas taitungensis* NC_009618. (2) Transition species between gymnosperm and angiosperms *Amborella trichopoda* NC_005086. (3) *Platanus occidentalis* NC_008335. (4) *Vitis vinifera* NC_007957. (5) *Nuphar advena* NC_008788. (6) *Nymphaea alba* NC_006050. (7) *Liriodendron tulipifera* NC_008326. (8) *Morus indica* NC_008359. (9) *Prunus persica* NC_014697. (10) *Populus trichocarpa* NC_009143. (11) *Populus alba* NC_008235. (12) *Arabidopsis thaliana* NC_000932. (13) Unicellular green alga *Euglena gracilis* NC_001603. (14) *Zea mays* NC_001666. (15) *Hordeum vulgare* NC_008590. (16) *Triticum aestivum* NC_002762. (17) *Oryza sativa Japonica* NC_001320. (18) *Oryza sativa indica* NC_008155. (19) *Equisetum arvense* NC_014699. (20) *Marchantia polymorpha* NC_001319. (21) *Lathyrus sativus* NC_014063. (22) *Welwitschia mirabilis* NC_010654. (23) *Cedrus deodara* NC_014575. (24) The longest-living plant *Pinus longaeva* NC_011157. (25) *Durinskia baltica* NC_014287. (26) *Pinus monophylla* NC_011158. (27) *Gnetum parvifolium* NC_011942. (28) *Ephedra equisetina* NC_011954. (29) *Cathaya argyrophylla* NC_014589 (described in 1955). (30) *Cuscuta obtusiflora* NC_009949. (31) *Euglena longa* NC_002652. (32) *Epifagus virginiana* NC_001568.

the molecular phylogeny and taxonomy (Ali et al., 2013). Recently ITS2 loci have been suggested as a universal DNA barcode for identification of plant species and as a complementary locus for CO1 to identify animal species (Chen et al., 2010; Yao et al., 2010). The ITS2 region has been shown to be applicable in discrimination among a wide range of plants within families of Asteraceae, Rutaceae, Rosaceae and so on (Gao et al., 2010; Luo et al., 2010; Yao et al., 2010; Pang et al., 2011). The analyses of ITS2 secondary structures of ribosomal genes have been used to improve the quality of phylogenetic reconstructions (Keller et al., 2010); however, its assessment for phylogeny at higher hierarchy is still lacking.

2.2. Chloroplast DNA (cpDNA)

Chloroplast DNA (Figs. 5 and 6) has been used very frequently in plant systematic and phylogenetic studies. It is a circular molecule ranging in size of 120–217 kb, with a unique exception of green alga *Floydiella terrestris* with huge cpDNA of 521.168 bp (NCBI # NC_014346) (Gyulai et al., 2012). There are about 100 functional genes in the chloroplast genome. It contains, with few exceptions (IRL – IRless), two duplicate regions in reverse orientation, known as the inverted repeats (IR) of 10–76 kb, which divides the chloroplast genome into large (LSC) and small single-copy (SSC) regions. The structural organization of chloroplast genome is highly conserved, i.e., relatively free of large deletions, insertions, transpositions, inversions and SNPs (single nucleotide polymorphism), which make it advantageous for phylogenetic studies. Chloroplast DNA is a relatively abundant (generally, 50 chloroplast per cell multiplied with 50 cpDNA copy per chloroplast) compared to nuclear DNA (generally 2n), which facilitating DNA extraction

and analysis. Chloroplast DNA is usually uniparentally inherited (maternally in angiosperms and paternally in gymnosperms, in general, with exceptions), which facilitates to determine the maternal parent in hybrids and allopolyploids (Ackerfield and Wen, 2003). Some chloroplast regions like *psbA-trnH* spacer, and *rps16* intron gene evolve relatively rapidly. There are a number of noncoding cpDNA regions which are also useful target of study such as the intergenic spacer of *atpB-rbcL* (Baker et al., 1999; Asmussen and Chase, 2001; Manen and Natali, 1995; Manen et al., 2002), the *rps16* intron, *matK*, *ndhF*, *ycf6-psbM*, and *psbM-trnD* (Oxelman et al., 1997; Andersson and Rova, 1999; Downie and Katz-Downie, 1999; Wallander and Albert, 2000; Štorchová and Olson, 2007), *rpL16* intron (Jordan et al., 1996; Baum et al., 1998), *trnL-F* (Wallander and Albert, 2000), and *psbA-trnH* spacer, *trnH-psbA* (Kress et al., 2005; Chase et al., 2005) by using universal primers.

For phylogenetic investigations cpDNA has been more readily exploited (Fig. 6), than the nuclear genome for barcoding, similar to mitochondrial genomes of animals. Kress et al. (2005) have compared plastid genomes of *Atropa* and *Nicotiana*, and recorded that nine intergenic spacers *trnK-rps16*, *trnH-psbA*, *rp136-rps8*, *atpB-rbcL*, *ycf6-psbM*, *trnV-atpE*, *trnC-ycf6*, *psbM-trnD*, and *trnL-F* fulfill the barcode criteria. For comparison, ITS had a much higher divergence value (13.6%) than any plastid regions, especially *rbcL*, which is far the lowest in divergence (0.83%).

2.3. Mitochondrial DNA (mtDNA)

Mitochondrial DNA (mtDNA) analysis has had a major impact on the study of phylogeny and population genetics in animals. Plant mtDNAs are rather poorly studied compared

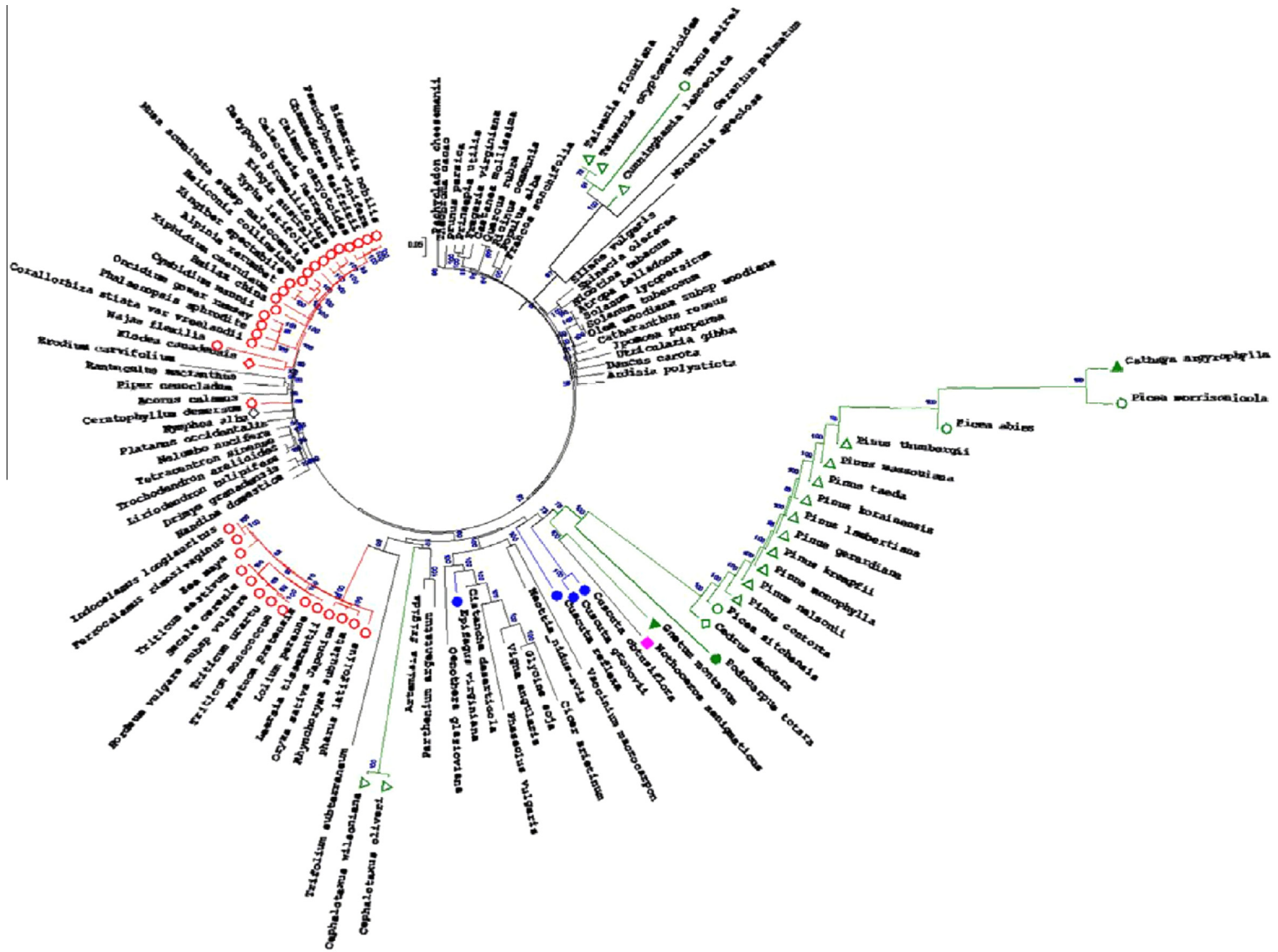


Figure 6 Cladogram of total cpDNAs. ML (Maximum Likelihood; Hillis et al., 1994) cladogram of total cpDNA genomes with 1000× bootstrap values (MEGA4; Tamura et al., 2007). Polyphyletic Dicots of Angiosperm are not labeled; monocots (○) show four cpDNA lineages including the individual *Acorus*. Gymnosperm clades (green) show three lineages: (1) *Gnetum* (▼)–*Podocarpus* (●) and the Pinaceae species of *Pinus*, *Picea*, *Abies*, *Cathaya* and *Cedrus*. *Cathaya* (▲) (NC_014589) has the smallest cpDNA genome (107.122 bp) of Coniferales. This clade shows close genetic distance to the bryophyte *Nothoceros*; (2) Taxodiaceae including *Taiwania*, *Taxus*, and *Cunninghamia*, which clade is close to eudicot *Monsomia* and *Geranium*; and (3) the evolutionarily youngest gymnosperm Cephalotaxaceae (*Cephalotaxus*) of Coniferales. Water submerged plants of monocot *Elodea* and dicot *Ceratophyllum* (◇); and parasite plants with reduced cpDNA genomes of *Cuscuta* and *Epifagus* (●) are also labeled. Scale (0.05) shows relative genetic distances based on substitution rate. Accession numbers are available at NCBI (<http://www.ncbi.nlm.nih.gov>) (Altschul et al., 1997) and CGP (<http://chloroplast.ocean.washington.edu>). For computing about 21 million nucleotides an eight core computer with 24 GB RAM was used running for 2–3 weeks.

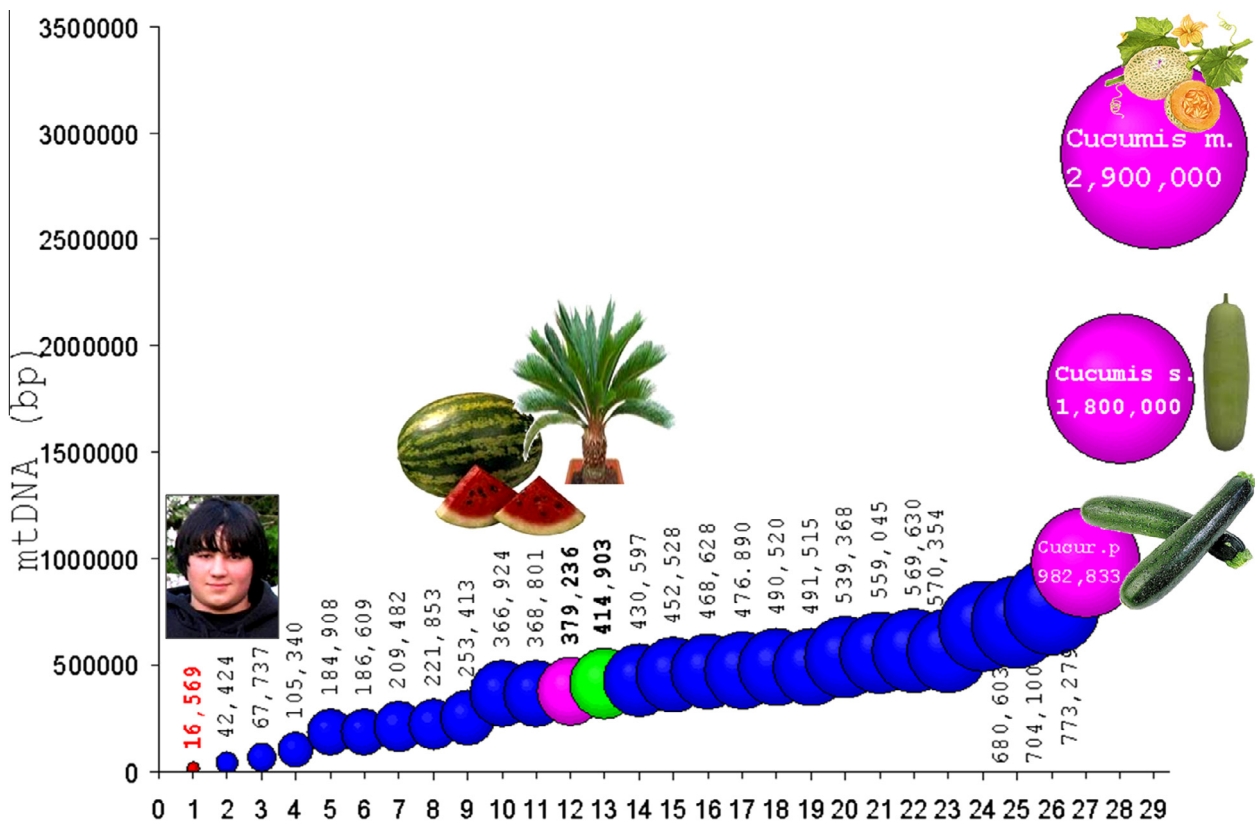


Figure 7 The mtDNA (*mitome*) sizes (bp) of organisms in increasing order including the only wood available *Cycas* (13). Human mtDNS (1), and lower plants (2–7) are indicated (NCBI accession #, [Altschul et al. 1997](#)). (1) *Homo s.* NC_012920. (2) *Mesostigma v.* NC_008240. (3) *Chara v.* NC_005255. (4) *Physcomitrella p.* NC_007945. (5) *Megaceros ae.* NC_012651. (6) *Marchantia p.* NC_001660. (7) *Phaeoceros l.* NC_013765. (8) *Brassica n.* NC_008285. (9) *Silene l.* NC_014487. (10) *Arabidopsis th.* NC_001284. (11) *Beta v.* NC_002511. (12) *Citrullus l.* NC_014043. (13) *Cycas taitungensis* NC_010303. (14) *Nicotiana t.* NC_006581. (15) *Triticum ae.* NC_007579. (16) *Sorghum b.* NC_008360. (17) *Carica p.* NC_012116. (18) *Oryza s. J.* NC_011033. (19) *Oryza s. I.* NC_007886. (20) *Zea lux.* NC_008333. (21) *Oryza r.* NC_013816. (22) *Zea m.* NC_007982. (23) *Zea pren.* NC_008331. (24) *Zea parv.* NC_008332. (25) *Tripsacum d.* NC_008362. (26) *Vitis v.* NC_012119. (27) *Cucurbita p.* NC_014050. (28) *Cucumis sativus.* (29) *Cucumis melo* with the largest mtDNA. The mtDNAs of Human (1), Cucurbits (12, 27, 28, 29), and gymnosperm *Cycas* (13) are indicated.

to animals ([Palmer et al., 2000](#)). Plant mtDNA encodes approximately 5% of the proteins found in the mitochondrion. Plant mtDNA is ‘abnormally’ large and variable in size (300–600 kb), many times larger than animal mtDNAs (16–25 kb). In Cucurbitaceae and Malvaceae, the mitochondrial genome was reported to exceed 2900 kb (*Cucumis melo*) ([Alverson et al., 2010](#)) (Figs. 7, 10 and 11). Many foreign sequences are found in plant mtDNA. Chloroplast DNA sequences as large as 12 kb in length, are found integrated in plant mtDNAs, and cpDNA sequences can fill up about 5–10% of the mtDNA. Plant mitochondrial genome seems to be unstable for barcoding for its frequent intramolecular and intermolecular recombinations, which continuously change gene orders ([Palmer, 1992](#); [Palmer et al., 2000](#)). However, nucleotide substitution rates of mtDNA are 3–4 times lower than cpDNA, about 12 times lower than those of plant nuDNA, and 40–100 times lower than those of animal mtDNA ([Cho et al., 2004](#)). Only a few mitochondrial markers have shown promise for phylogenetic utility ([Demesure et al., 1995](#); [Freudenstein and Chase, 2001](#)). However, the divergence of coding region of cytochrome oxidase subunit 1 (*cox1*) (referred as CO1) among plant and animal families has found documented to be only

a few base pairs across 1.4 kb of sequence ([Hebert et al., 2003](#); [Folmer et al., 1994](#)), which makes it useful for barcoding.

3. Molecular phylogenetic analyses

In DNA barcoding the sequences of the barcoding region are obtained from various individuals. The resulting sequence data are then used to construct a phylogenetic tree. In such a tree, similar, putatively related individuals are clustered together. The term ‘DNA barcode’ seems to imply that each species is characterized by a unique sequence, but there is of course considerable genetic variation within each species as well as between species. However, genetic distances between species are usually greater than those within species, so the phylogenetic tree is characterized by clusters of closely related individuals, and each cluster is assumed to represent a separate species ([Dasmahapatra and Mallet, 2006](#)).

GenBank (-the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences) has a very important role in DNA barcoding. GenBank is part of the International Nucleotide Sequence Database Collaboration, which

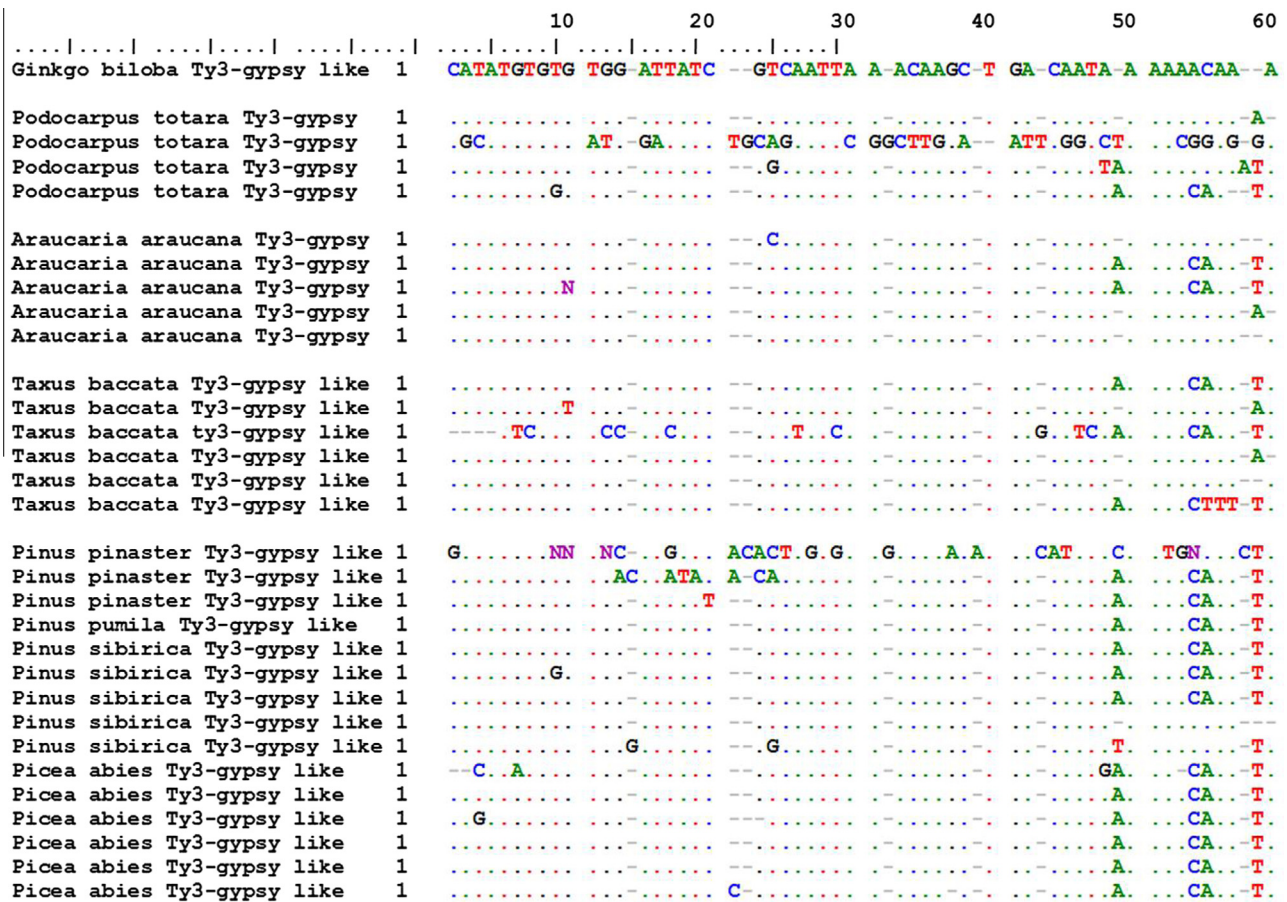


Figure 8 Samples' (60 nt) sequence polymorphism of *pPongy2* LTR retrotransposons. Spread and sequence diversity of RT (*reverse transcriptase*) gene of *pPongy2*, a Ty3-gypsy-like LTR-retrotransposon through the evolution of gymnosperms from *Ginkgo* → *Podocarpus* (248 My) → *Araucaria* (230 My) → *Taxus* (206 My) → *Pinaceae* (180 My). Sequence # AJ290647.1 was downloaded from NCBI and aligned by BioEdit program (Hall, 1999).

comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. The GenBank sequence database is an open access, annotated collection of all publicly available nucleotide sequences and their protein translations. This database is produced at National Center for Biotechnology Information (NCBI) as part of the International Nucleotide Sequence Database Collaboration. GenBank and its collaborators receive sequences produced in laboratories throughout the world from more than 100,000 distinct organisms. GenBank is built by direct submissions from individual laboratories, as well as from bulk submissions from large-scale sequencing centers. GenBank has merged the values of natural history with those of the experimental sciences. The Entrez Nucleotide and BLAST (Basic Local Alignment Search Tool) are the two main ways to search and retrieve data from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>).

A wide range of programs (detailed information can be browsed at <http://bioinformatics.unc.edu/software/opensource/index.htm>, http://molbiol-tools.ca/molecular_biology_freeware.htm#Phylogeny, and <http://evolution.genetics.washington.edu/phylip/software.html#recent>) are available for sequence data analysis. The three commonly used methods for phylogenetic analysis are MP (maximum parsimony), ML (maximum

likelihood), and (BI) Bayesian inference. Of them ML (maximum likelihood) was found to be the most discriminative (Hillis et al., 1994). The maximum parsimony algorithm (Farris, 1970; Swofford et al., 1996) searches for the minimum number of genetic events (*e.g.* nucleotide substitutions) to infer the shortest possible tree (*i.e.*, the maximally parsimonious tree). Often the analysis generates multiple equally most parsimonious trees. When evolutionary rates are drastically different among the species analyzed, results from parsimony analysis can be misleading (*e.g.*, long-branch attraction; Felsenstein, 1978). Parsimony analysis is most often performed with the computer program PAUP* 4.0 (Swofford, 2002), and MEGA (Tamura et al., 2007, 2011). The maximum likelihood (ML) method (Felsenstein, 1985; Hillis et al., 1994) evaluates an evolutionary hypothesis in terms of the probability that the proposed model and the hypothesized history would give rise to the observed data set properly. The topology with the highest maximum probability or likelihood is then chosen. This method may have lower variance than other methods and is thus least affected by sampling error and differential rates of evolution. It can statistically evaluate different tree topologies and use all of the sequence information. The Bayesian phylogenetic inference is model-based method and was proposed as an alternative to maximum likelihood (Rannala and Yang, 1996; Yang and Rannala,



Figure 9 The *pPongy2* cladogram. Fast Minimum Dendrogram edited by NCBI server (Altschul et al., 1997) shows the spread among- and-within species of RT (reverse transcriptase) gene of *pPongy2*, a Ty3-gypsy-like LTR-retrotransposon through the evolutionary lineage of gymnosperms from *Ginkgo* → *Podocarpus* (248 My) → *Araucaria* (230 My) → *Taxus* (206 My) → *Pinaceae* (180 My). Genetic distance (scale 0.005) and branch length are indicated, and gymnosperm species are labeled with different color symbols. The accession numbers of taxon included in analyses were *Araucaria araucana* (AJ290651, AJ290652, AJ290653, AJ290654, AJ290655), *Ginkgo biloba* (AJ290656), *Picea abies* (AJ290585, AJ290586, AJ290591, AJ290592, AJ290593, AJ290594), *Pinus pinaster* (AJ290605, AJ290606), *Pinus pumila* (AJ290616), *Pinus sibirica* (AJ290623, AJ290626, AJ290629, AJ290630, AJ290631), *Podocarpus totara* (AJ290647, AJ290648, AJ290649, AJ290650), *Taxus baccata* (AJ290640, AJ290641, AJ290642, AJ290643, AJ290644, AJ290645).

1997). The computer program MrBayes 3.0 (Huelsenbeck and Ronquist, 2001) performs Bayesian estimation of phylogeny based on the posterior probability distribution of trees, which is approximated using a simulation technique called Markov chain Monte Carlo (or MCMC). MrBayes can combine information from different data partitions or subsets evolving under different stochastic evolutionary models. This allows the user to analyze heterogeneous data sets consisting of different data

types, including morphology and nucleotides. Bayesian inference has facilitated the exploration of parameter-rich evolutionary models (Table 2).

4. Retrotransposon based barcoding

Retrotransposons (RTs) are ubiquitous and dispersed throughout the host genome of an organism with correlations

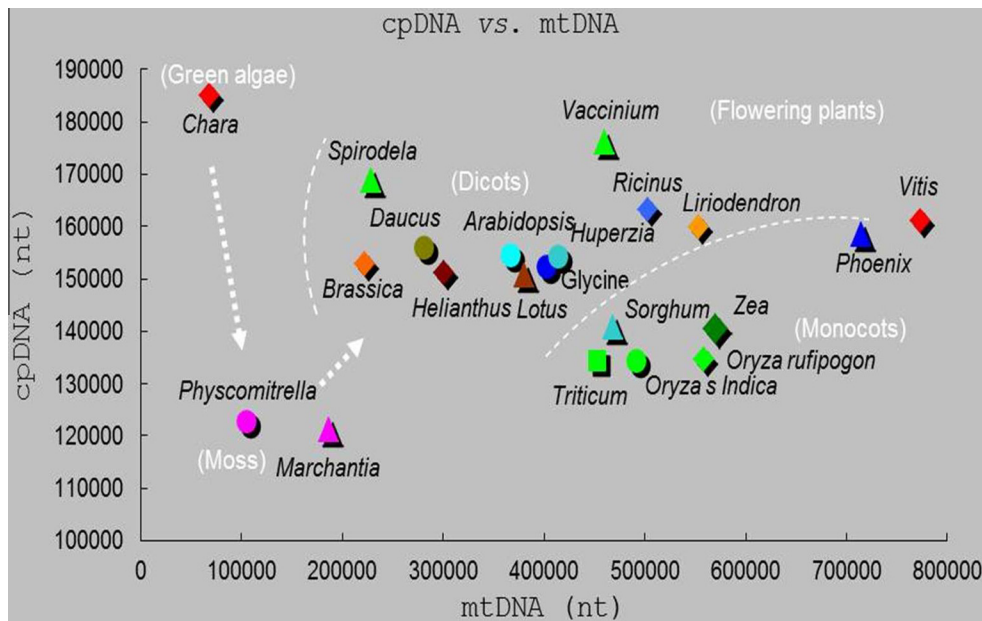


Figure 10 Size correlations between cpDNA and mtDNA genomes show a shift from green algae (*Chara vulgaris*) with high cpDNA/mtDNA ratio (2.73) through mosses of *Physcomitrella patens* (1.16) and *Marchantia polymorpha* (0.65) toward flowering plants of dicots to monocots with exception of *Vitis*. The decreasing ratio of cpDNA/mtDNA indicates an enlarging mtDNA during the evolution: *Spirodela polyrhiza* (0.74); *Brassica napus* (0.69); *Daucus carota* (0.55); *Helianthus annuus* (0.50); *Arabidopsis thaliana* (0.42); *Lotus japonicus* (0.40); *Vaccinium macrocarpon* (0.38); *Glycine max* (0.38); *Vigna radiata* (0.38); *Huperzia lucidula* (0.37); *Ricinus communis* (0.32); *Sorghum bicolor* (0.30); *Triticum aestivum* (0.29); *Liriodendron tulipifera* (0.29); *Oryza sativa Japonica* (0.274); *Oryza sativa Indica* (0.273); *Zea mays* (0.25); *Oryza rufipogon* (0.24); *Phoenix dactylifera* (0.22); *Vitis vinifera* (0.21). NCBI (Altschul et al., 1997) data were plotted by XY plot of Microsoft Windows Xcel program.

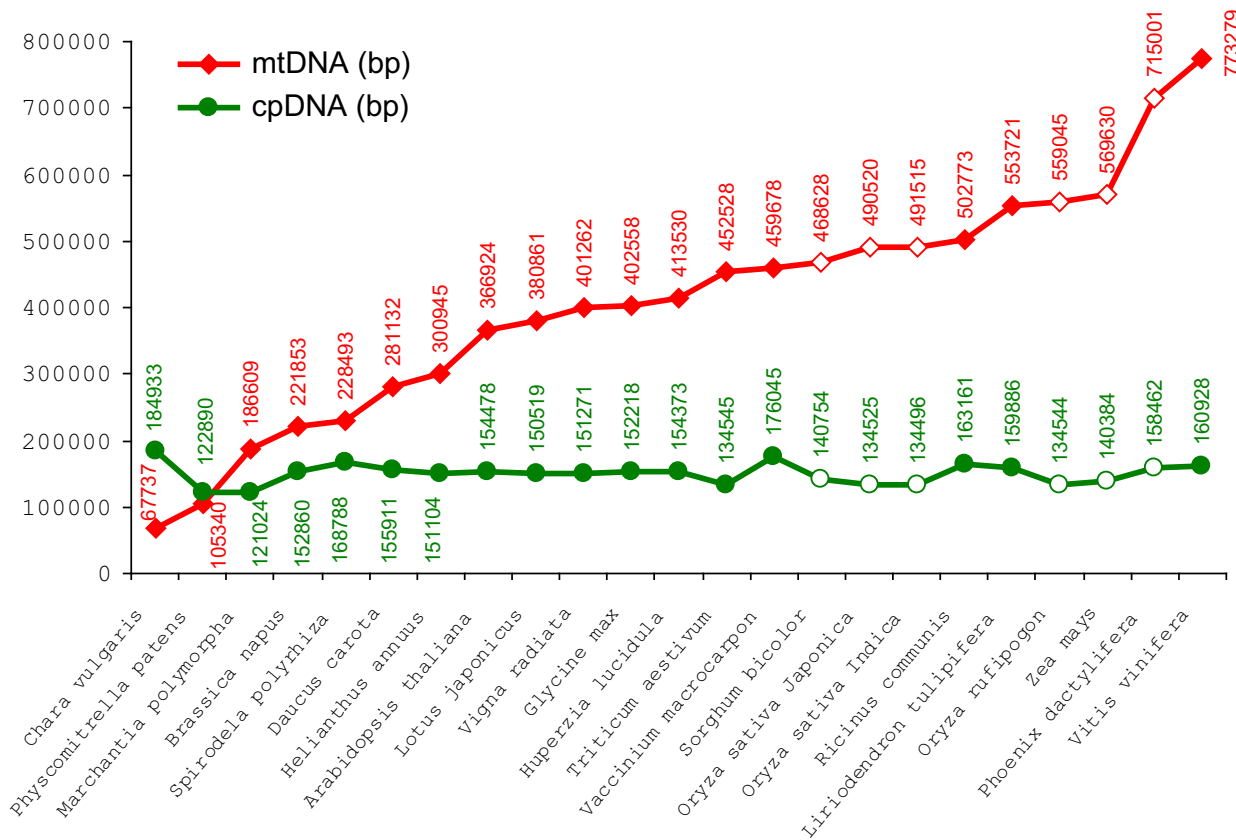


Figure 11 Changes of organelle genome sizes during the evolution (mtDNA – mitochondrial DNA; cpDNA chloroplast DNA) (see Fig. 10). NCBI (Altschul et al., 1997) data were plotted by Microsoft Windows Xcel program. Monocots are labeled with open symbols.

Table 2 A brief of the commonly used programs used for the phylogeny and DNA barcoding.

Phylogenetic analysis software	Description	Link/References
Bayesian evolutionary analysis sampling trees (BEAST)	A Bayesian MCMC program for inferring rooted trees under the clock or relaxed-clock models. It can be used to analyze nucleotide and amino acid sequences, as well as morphological data. A suite of programs, such as Tracer and FigTree, are also provided to diagnose, summarize and visualize results	http://beast.bio.ed.ac.uk Drummond and Rambaut (2007)
BioEdit	BioEdit is a fairly comprehensive sequence alignment and analysis tool. BioEdit supports a wide array of file types and offers a simple interface for local BLAST searches	http://www.mbio.ncsu.edu/bioedit/bioedit.html Hall (1999)
ClustalX	ClustalX is a windows interface for the ClustalW multiple sequence alignment program. It provides an integrated environment for performing multiple sequence and profile alignments and analyzing the results. This program allows to create Neighbor Joining trees with bootstrapping	http://www.clustal.org/ Thompson et al. (1997)
ClustalW	Multiple Sequence Alignment (EBI, United Kingdom). This provides one with a number of options for data presentation, homology matrices [BLOSUM (Henikoff), PAM (Dayhoff) or GONNET, and presentation of phylogenetic trees (Neighbor-Joining, Phylip or Distance). Other sites offering ClustalW alignment are at the Pasteur Institute, Kyoto University and chEMBLnet.org	http://www.ebi.ac.uk/Tools/msa/clustalw2/
DNA for Windows	DNA for Windows is a compact, easy to use DNA analysis program, ideal for small-scale sequencing projects	http://www.dna-software.co.uk/
Geneious	Geneious (Alexei Drummond Biomatters Ltd. Auckland, New Zealand) provides an automatically-updating library of genomic and genetic data; for organizing and visualizing data. It provides a fully integrated, visually-advanced toolset for: sequence alignment and phylogenetics; sequence analysis including BLAST; protein structure viewing, NCBI, EMBL, Pubmed auto-find, etc.	http://www.geneious.com/
MAFFT	MAFFT is a multiple sequence alignment program for unix-like operating systems. It offers a range of multiple alignment methods, L-INS-i (accurate; for alignment of $< \sim 200$ sequences), FFT-NS-2 (fast; for alignment of $< \sim 10,000$ sequences), etc.	http://mafft.cbrc.jp/alignment/software/
FigTree	FigTree is designed as a graphical viewer of phylogenetic trees to display summarized and annotated trees produced by BEAST	http://tree.bio.ed.ac.uk/software/figtree/
Format Converter v2.2.5	This program takes as input a sequence or sequences (<i>e.g.</i> , an alignment) in an unspecified format and converts the sequence(s) to a different user-specified format	http://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html#details_section http://code.google.com/p/garli Zwickl (2006)
Genetic algorithm for rapid likelihood inference (GARLI)	A program that uses genetic algorithms to search for maximum likelihood trees. It includes the GTR + Γ model and special cases and can analyze nucleotide, amino acid and codon sequences. A parallel version is also available	http://www.hyphy.org Kosakovsky et al. (2005)
Hypothesis testing using phylogenies (HYPHY)	A maximum likelihood program for fitting models of molecular evolution. It implements a high-level language that the user can use to specify models and to set up likelihood ratio tests	http://its2.bioapps.biozentrum.uni-wuerzburg.de/
ITS2 Database	The ITS2 Database presents an exhaustive dataset of internal transcribed spacer 2 sequences from NCBI GenBank accurately reannotated. Following an annotation by profile Hidden Markov Models (HMMs), the secondary structure of each sequence is predicted. The ITS2 Database also provides several tools to process your own ITS2 sequences, including annotation, structural prediction, motif detection and BLAST (Altschul and Gapped, 1997) search on the combined sequence–structure information. Moreover, it integrates trimmed versions of 4SALE (Seibel et al., 2006, 2008) and ProfDistS (Wolf et al., 2008) for multiple sequence–structure alignment calculation and Neighbor Joining (Saitou and Nei, 1987) tree reconstruction. Together they form a coherent analysis pipeline from an initial set of sequences to a phylogeny based on sequence and secondary structure	
Molecular evolutionary genetic analysis (MEGA)	A Windows-based program with a full graphical user interface that can be run under Mac OSX or Linux using Windows emulators. It includes distance, parsimony and likelihood methods of phylogeny reconstruction, although its strength lies in the distance methods. It incorporates the alignment program ClustalW and can retrieve data from GenBank	http://www.megasoftware.net Tamura et al. (2011)

(continued on next page)

Table 2 (continued)

Phylogenetic analysis software	Description	Link/References
MrBayes	A Bayesian MCMC program for phylogenetic inference. It includes all of the models of nucleotide, amino acid and codon substitution developed for likelihood analysis	http://mrbayes.net Huelsenbeck and Ronquist (2001)
Modeltest	Modeltest is a program that uses hierarchical likelihood ratio tests (hLRT) to compare the fit of the nested GTR (General Time Reversible) family of nucleotide substitution models. Additionally, it calculates the Akaike Information Criterion estimate associated with the likelihood scores	http://darwin.uvigo.es/software/jmodeltest.html Posada and Crandall (1998)
Oligo Calculator	On line tool for to find length, melting temperature, %GC content and molecular weight of DNA sequence	http://mbcf.dfcf.harvard.edu/docs/oligo.html
Phylogenetic analysis by maximum likelihood (PAML)	A collection of programs for estimating parameters and testing hypotheses using likelihood. It is mostly used for tests of positive selection, ancestral reconstruction and molecular clock dating. It is not appropriate for tree searches	http://abacus.gene.ucl.ac.uk/software Yang (2007)
Phylogeny.fr	Phylogeny.fr – is a simple to use web service dedicated to reconstructing and analyzing phylogenetic relationships between molecular sequences. It includes multiple alignment (MUSCLE, T-Coffee, ClustalW, ProbCons), phylogeny (PhyML, MrBayes, TNT, BioNJ), tree viewer (Drawgram, Drawtree, ATV) and utility programs (<i>e.g.</i> Gblocks to eliminate poorly aligned positions and divergent regions)	http://www.phylogeny.fr/ Dereeper et al. (2008)
PHYLIP	A package of programs for phylogenetic inference by distance, parsimony and likelihood methods	http://evolution.gs.washington.edu/phylip.html
PhyML	A fast program for searching for the maximum likelihood trees using nucleotide or protein sequence data	http://www.atgc-montpellier.fr/phyml/binaries.php Guindon and Gascuel (2003)
PAUP	David Swofford of the School of Computational Science and Information Technology, Florida State University, Tallahassee, Florida has written PAUP* (which originally meant Phylogenetic Analysis Using Parsimony). PAUP* version 4.0beta10 has been released as a provisional version by Sinauer Associates, of Sunderland, Massachusetts. It has Macintosh, PowerMac, Windows, and Unix/OpenVMS versions. PAUP* has many options and close compatibility with MacClade. It includes parsimony, distance matrix, invariants, and maximum likelihood methods and many indices and statistical tests	http://paup.csit.fsu.edu Swofford (2002)
ProfDistS	Distance based phylogeny on sequence–structure alignments. Bioinformatics. 24, 2401–2402	Wolf et al. (2008)
MacClade	MacClade is a computer program for phylogenetic analysis written by David Maddison and Wayne Maddison. Its analytical strength is in studies of character evolution. It also provides many tools for entering and editing data and phylogenies, and for producing tree diagrams and charts	http://macclade.org/
Neighbor-Joining	Neighbor-Joining method is proposed for reconstructing phylogenetic trees from evolutionary distance data	Saitou and Nei (1987)
PHYLIP	PHYLIP (the PHYLogeny Inference Package) is a package of programs for inferring phylogenies. PHYLIP is the most widely-distributed phylogeny package, and competes with PAUP to be the one responsible for the largest number of published trees	http://evolution.genetics.washington.edu/phylip.html
RAxML	A fast program for searching for the maximum likelihood trees under the GTR model using nucleotide or amino acid sequences. The parallel versions are particularly powerful	http://scof-its.org/exelixis/software.html Stamatakis (2006)
Readseq	A tool for converting between common sequence file formats, particularly useful for those using various phylogenetic analysis tools	http://www.ebi.ac.uk/Tools/sfc/readseq/
4SALE	A tool for synchronous RNA sequence and secondary structure alignment and editing	Seibel et al. (2006, 2008)
Sequencher	The Premier DNA Sequence Analysis Software for Sanger and NGS Datasets	http://www.genecodes.com/
Tree analysis using new technology (TNT)	A fast parsimony program intended for very large data sets	http://www.zmuc.dk/public/phylogeny/TNT Goloboff et al. (2008)
TreeView	TreeView provides a simple way to view the contents of a NEXUS, PHYLIP, or other format tree file	http://taxonomy.zoology.gla.ac.uk/rod/treeview.html

to genome sizes (Alzohairy et al., 2012, 2013). RTs are classified into two main groups of long terminal repeats (LTR-RTs) and non-LTR retrotransposons with several subgroups (*i.e.* families). The copy-and-paste way of transposition of RTs consists of three molecular steps including transcription of an RNA copy from the genomic RT, followed by reverse transcription of the RNA copy to cDNA, which is synthesized by reverse transcriptase (RTase), and the final reintegration of cDNA-RT into a new location of the genome. These processes lead to new and new genomic insertions of the RTs without excision of the original element. Finally, a population of the given RT develops hundreds and hundreds of copies in the host genome with sequence modifications (*i.e.* polymorphism) due to the lack of proofreading activity of RTases (Kalendar et al., 2000). All these molecular characteristics of RTs provide targets for barcoding (Lágler et al., 2006; Schulman, 2007; Mansour et al., 2009a,b), and tracking of horizontal and vertical spread of RTs (Figs. 8 and 9).

5. Metabarcoding

In the early days, DNA barcoding mainly focused on taxonomic research. Nowadays, DNA barcoding could be used for a wide variety of purposes and the research field has diversified itself. Not only DNA of intact and isolated species is used, but also DNA that is shed into the environment by organisms (*e.g.* DNA from skin, nails, hairs, waste products, etc.). This form of DNA is called environmental DNA, or *eDNA*, and is usually highly degraded. It could be found in environmental samples like air, water or soil and it could be extracted without isolating the organisms (Taberlet et al., 2012a). The eDNA fragments are shorter than normal DNA fragments and thus an adjusted barcoding method is needed. Specifically, it requires the use of shorter barcodes (Hajibabaei et al., 2006). A second direction of development concerns the use of environmental samples, not only to identify one single species, but to identify a wide variety of species in one experiment. This is called metabarcoding. The development of metabarcoding approaches was aided by the discovery of NGS, which allows parallel reading of DNA sequences from a single DNA-extract without a necessity for cloning (*i.e.* Taberlet et al., 2012a,b,c; Hajibabaei, 2012). In this respect, metabarcoding differs from normal DNA barcoding in the sense that classic DNA barcoding aims to identify intact *specimens* (*e.g.* with complete genomes) up to species level, and metabarcoding aims to identify degraded DNA samples (eDNA) up to family level or higher.

6. Materials for DNA barcoding

DNA is a relatively stable molecule, besides fresh sample, DNA can also be isolated from museum collections, including animal specimens preserved in formalin (Fang et al., 2002). Plant DNA can be extracted from herbarium specimens up to 100 years old, and also from archeological plant remains (Szabó et al., 2005; Lágler et al., 2005; Gyulai et al., 2006, 2012; Palmer et al., 2012). Plant DNA quality depends on methodology adopted for drying after being pressed. If specimen-drying facilities are not immediately available, especially in humid tropical climates, botanists often treat pressed samples with ethanol to

temporarily preserve against fungal attack and degradation. Alcohol has been shown to be detrimental to recovering high-quality DNA. It is encouraging that museum specimens of insects dried from ethanol storage readily yield CO1 sequences.

7. Benefits of DNA barcoding

DNA barcoding is of great utility to users of taxonomy. It provides more rapid progress than the traditional taxonomic work (Gregory, 2005). DNA barcoding allows taxonomists to rapidly sort specimens by highlighting divergent taxa that may represent new species. DNA barcoding offers taxonomists the opportunity to greatly expand, and eventually complete, a global inventory of life's diversity. The advocates of DNA barcoding say that it has revitalizing biological collections and speed up species identification and inventories (Gregory, 2005; Schindel and Miller, 2005); however the opponents argue that it will destroy traditional systematics and turn it into a service industry (Ebach and Holdrege, 2005; Seberg et al., 2003). Once fully developed, DNA barcoding will have the potential to completely revolutionize our knowledge of diversity of living organisms and our relationship to nature. By harnessing technological advances in electronics and genetics, DNA barcoding will help many people to quickly and cheaply recognize known species and retrieve information about them, and will speed discovery of thousands of species yet to be named. Barcoding has the potential to provide a vital new tool for appreciating and managing the Earth's immense biodiversity.

The urgency of creating tissue banks has been well recognized (Savolainen and Reeves, 2004; Lorenz et al., 2005), and solutions for linking DNA samples with taxonomic vouchers are being developed for all sorts of organisms. Barcoding of life will have to be both integrative and integrated with other taxonomic initiatives such as the Global Taxonomic Initiative of the Convention on Biological Diversity (www.biodiv.org), and the Global Biodiversity Information Facility (www.gbif.org). Finally, by barcoding of life, 'Life Barcoders' will identify species linked via the World Wide Web to other kinds of biodiversity data such as images, usage and conservation status.

Despite some drawbacks of using DNA barcoding, the reported success of using the barcoding region in distinguishing species from a range of taxa and to reveal cryptic species is remarkable. However, it is known that species identification based on a single DNA sequence will always produce some erroneous results. Efforts should therefore be made to develop nuclear barcodes to complement the barcoding region that is currently in use. As the advantages and limitations of barcoding become apparent, it is clear that taxonomic approaches integrating DNA sequencing, morphology and ecological studies will achieve maximum efficiency at species identification (Dasmahapatra and Mallet, 2006).

The main reasons of DNA barcoding are (a) DNA barcode works with fragments, (b) DNA barcode works for all stages of life, (c) DNA barcoding unmasks those species which look-alikes, (d) DNA barcoding reduces ambiguity, (e) DNA barcoding makes expertise to speed up the identification of known organisms and also facilitate rapid recognition of new species, (f) DNA barcoding facilitates democratizes access *i.e.* a standardized library of barcodes will empower many more people to call by name the species around them as well as it will make possible identification of species whether

abundant or rare, native or invasive, engendering appreciation of biodiversity locally and globally, (g) DNA barcoding sprouts new leaves on the tree of life, (h) DNA barcoding demonstrates value of specimens collections, (i) DNA barcoding speeds writing the encyclopedia of life, and (j) barcoding links biological identification to advancing frontiers in DNA sequencing, miniaturization in electronics, computerized information storage, and these integrating will lead to portable desktop devices and ultimately to hand-held barcoders; however, promoting technology development of portable devices for field use will be a major goal of this initiative (<http://barcoding.si.edu/PDF/TenReasonsBarcoding.pdf>).

Although DNA sequence data and barcoding are well on the way to being accepted as the global standard for species identification; however, such development is still limited in use. With the rich biological resources in many developing countries and many excellent taxonomists who are intimately familiar with the regional flora and interesting systematic questions, more plant DNA barcoding and molecular systematic studies by colleagues from developing countries should advance our understanding of the tree of life at the global scale and offer opportunities to address many new evolutionary questions as well. Plant DNA barcoding and molecular systematic research require more equipment for data collection and analysis. It is technically more expensive than the classical, morphological and anatomical studies, but perhaps affordable. There is need to harness the mountains of DNA data being generated in modern laboratories and also to use the data from deep morphology in systematic (Ali and Choudhary, 2011; Wen and Pandey, 2005).

7.1. Plant systematics

DNA sequence data and barcoding are well on the way to being accepted as the global standard for species identification (Ali and Choudhary, 2011). DNA barcodes are likely to play a major role in the future of taxonomy. The build-up of DNA databases has great potential for the identification and classification of organisms and for supporting ecological and biodiversity research programs (Tautz et al., 2003). As a uniform, practical method for species identification, it appears to have broad scientific applications. DNA-based identification of species offers enormous potential for the biological scientific community, educators, and the interested general public. It will help open the treasury of biological knowledge and increase community interest in conservation biology and understanding of evolution.

The direct benefits of DNA barcoding are to make the outputs of systematics available to a large number of end-users by providing standardized and high-tech identification tools, e.g. for biomedicine (parasites and vectors), agriculture (pests), environmental assays and customs (trade in endangered species). The future perspective of DNA barcoding will be to provide a bio-literacy tool for the general public and it will also help in opening to the treasury of biological knowledge, which is currently underused partly because of the weak taxonomic expertise for species identification. DNA barcoding will also relieve the enormous burden of identifications of taxonomists, so they can focus on more pertinent to discovering and describing the new species. The most important aspect of DNA barcoding is that it will

facilitate basic biodiversity inventories (Savolainen et al., 2005; Lahaye et al., 2008).

DNA barcoding can be likened to aerial photography, in that it provides an efficient method for mapping the extant species, though in sample space rather than physical space. The “aerial map” of DNA barcodes will help investigators explore the biological world and make full use of the enormous knowledge that has been built on 250 years of classical taxonomy. As sequencing costs decrease, DNA-based species identification will become available to an increasingly wide scientific community. When costs are low enough, researchers, teachers and naturalists will be able to use DNA barcoding in depth for examination of local ecosystems. As DNA barcodes are applicable to all life stages, it is also useful in cases where e.g. larval stages are difficult to identify with traditional methods of butterflies (Janzen et al., 2005) or amphibians (Vences et al., 2005), and insects in which several casts have different ‘unrelated’ morphologies (Smith et al., 2005). However, DNA barcoding is applied only in conjunction with classical approaches based on morphology.

7.2. Medicinal and wild plants’ identification

One of the most important uses of the DNA barcoding is in the medicinal plant authentication. Recently ITS, *trnH-psbA*, *rbcL*, *matK* and *trnL-trnF* gene sequences have successfully been used for DNA barcoding of several plant species, such as ITS [*Achyranthes bidentata* (Wang et al., 2004), *Aconitum* species (Luo and Yang, 2008; Zhang et al., 2010b), *Adenophora lobophylla* (Ge et al., 1997), *Alpinia* (Zhao et al., 2000), *Alpinia galangal* (Zhao et al., 2001), *Alpinia oxyphylla* (Zhao et al., 2000), *Amomum* species (Pan et al., 2001; Zhou et al., 2002), *Angelica sinensis* (Ji et al., 2002; Zhang et al., 2003; Zhao et al., 2006), *Aquilaria sinensis* (Shen et al., 2008; Niu et al., 2010), *Arctium lappa* (Liu et al., 2010), *Astragalus* species (Dong et al., 2003), *Atractylodes* species (Shiba et al., 2006), *Bupleurum* species (Xie et al., 2006; Yang et al., 2007; Xie et al., 2009), *Changium smyrnioides* (Tao et al., 2008), *Chuanminshen violaceum* (Tao et al., 2008), *Citrus medica* var. *sarcodactylis* (Gao et al., 2007), *Cnidium monnieri* (Cai et al., 2000), *Codonopsis tangshen* (Luo et al., 2010), *Crocus sativus* (Mao et al., 2007; Che et al., 2007), *Cynanchum* species (Zhang et al., 2010a); *Dendrobium chrysanthum* (Xu et al., 2001), *Dendrobium nobile* (Ge et al., 2008), *Dendrobium officinale* (Ding et al., 2002a), *Dendrobium* species (Lau et al., 2001; Ding et al., 2002a,b,c; Xu et al., 2006), *Dioscorea* species (Wang et al., 2007), *Ephedra* species (Guo et al., 2006), *Eucommia ulmoides* (Ma et al., 2004), *Euphorbia* species (Jiang et al., 2005), *Gentiana dahurica* (Ji et al., 2003b), *Gynostemma pentaphyllum* (Jiang et al., 2009), *Hedyotis diffusa* (Hao et al., 2004; Liu and Hao, 2005), *Hypericum perforatum* (Howard et al., 2009), *Ligusticum chuanxiong* (Liu et al., 2002), *Liriope* species (Huang et al., 2009), *Lycium barbarum* (Shi et al., 2008), *Mitragyna speciosa* (Sukrong et al., 2007), *Morinda officinalis* (Ding and Fang, 2005), *Nelumbo nucifera* (Lin et al., 2007), *Ophiopogon japonicus* (Huang et al., 2009), *Panax ginseng* (Ma et al., 2000), *Panax* species (Ngan et al., 1999), *Polygonum multiflorum* (Zhang and Shi, 2007), *Polygonum tinctorium* (Song et al., 2009), *Pseudostellaria heterophylla* (Yu et al., 2003; Zhu et al., 2007), *Pueraria* species (Zeng et al., 2003; Sun et al., 2007), *Rheum palmatum* (Zhang et al., 2003; Ji et al., 2003a), *Rhodiola alsia* (Gao et al., 2009), *Salvia miltiorrhiza* (Wang and Wang, 2005), *Saussurea lappa* (Chen et al., 2008),

Saussurea medusa (Liu et al., 2001b), *Schisandra chinensis* (Gao et al., 2003), *Stellaria media* (Zhao et al., 2009), *Stemona tuberosa* (Jiang et al., 2006), *Swertia mussoitii* (Liu et al., 2001a), *Tripterygium wilfordii* (Law et al., 2010), *Verbena officinalis* (Ruzicka et al., 2009)], *trnH-psbA* [*Aristolochia* species (Li et al., 2010), *Artemisia* species (Liu and Ji, 2009), *Citrus grandis* (Su et al., 2010), *Dendrobium* species (Yao et al., 2009), *Paris* species (Yang et al., 2010), *Sabia parviflora* (Sui et al., 2010), Species in *Polygonaceae* (Song et al., 2009), *Stemona tuberosa* (Vongsak et al., 2008)], *rbcL* [*Arisaema* species (Kondo et al., 1998), *Belamcanda chinensis* (Qin et al., 2003), *Cnidium officinale* (Kondo et al., 1996), *Dendrobium* species (Asahina et al., 2010), *Dryopteris crassirhizoma* (Zhao et al., 2007), *Glycyrrhiza* species (Hayashi et al., 1998, 2000, 2005), *Sabia parviflora* (Sui et al., 2010), Species in *Polygonaceae* (Song et al., 2009)], *matK* [*Cnidium monnieri* (Cao et al., 2001), *Cnidium officinale* (Liu et al., 2002), *Dendrobium* species (Teng et al., 2002; Asahina et al., 2010), *Ligusticum chuanxiong* (Liu et al., 2002), *Panax notoginseng* (Fushimi et al., 2000; Zhang et al., 2006), *Panax* species (Zhu et al., 2003), *Panax vietnamensis* (Komatsu et al., 2001), *Polygonum multiflorum* (Yan et al., 2008), *Rheum* species (Yang et al., 2004), *Sabia parviflora* (Sui et al., 2010)], and *trnL-trnF* [*Adenophora* species (Zhao et al., 2003), *Angelica acutiloba* (Mizukami et al., 1997), *Angelica* species (Mizukami, 1995), *Atractylodes* species (Ge et al., 2007), *Cinnamomum* species (Kojoma et al., 2002), *Epimedium* species (Sun et al., 2004), *Fritillaria* species (Cai et al., 1999), *Lonicera japonica* (Li et al., 2001), *Pueraria* species (Sun et al., 2007), *Saussurea lappa* (Chen et al., 2008), *Stellaria media* (Zhao et al., 2009), *Swertia mussoitii* (Yu et al., 2008), *Tripterygium wilfordii* (Law et al., 2010)].

In addition with the above, Chen et al. (2010) tested the discrimination ability of ITS2 in more than 6600 plant samples belonging to 4800 species from 753 distinct genera (see the link for reference: <http://www.plosone.org/article/doi/10.1371/journal.pone.0008613.s008>) and found that the rate of successful identification with the ITS2 was 92.7% at the species level. Yao et al. (2010) also evaluated 50,790 plant and 12,221 animal (see the link for reference: <http://www.plosone.org/article/doi/10.1371/journal.pone.0013102.s006>) ITS2 sequences downloaded from GenBank, and proposed that the ITS2 locus should be used as a universal DNA barcode for identifying plant species and as a complementary locus for CO1 to identify animal species.

7.3. Food safety and conservation biology

Sustainable utilization of plant genetic resources is essential to meet the demand for future food and health security. Molecular markers are increasingly used for screening of germplasm to study genetic diversity, identify redundancies in the collections (Rao, 2004). Despite the tradition of systematic biology as the science of diversity, systematics has until recently contributed relatively little to the theory and practice of conservation biology. The four areas in which systematics could contribute to the conservation of rare plant species are: (i) species concepts, (ii) the identification of lineages worthy of conservation, (iii) the setting of conservation priorities, and (iv) the effects of hybridization on the biology and conservation of rare species. Species concepts that incorporate history and reflect phylogeny ultimately are more useful for preserving

biodiversity. Phylogenetic analyses involving conspecific populations often reveal multiple lineages that may warrant protection as evolutionarily distinct units. Phylogenetic information provides the tools for inferring relationships among organisms and, in conjunction with biogeography, for identifying those areas that harbor many actively speciation groups. Hybridization may lead to the extinction of a rare species, but in other cases, ironically, artificial hybridization with a more widespread congener may be the only way to preserve the gene pool of a rare species (Soltis and Gitzendanner, 1999).

A common problem with raw drug trade has been the admixtures with morphologically allied and geographically co-occurring species (Nair et al., 1983; Bisset, 1984; Sunita, 1992; Khattoon et al., 2006; Mitra and Kannan, 2007). Over 80% of the medicinal plants for raw drug trade are predominantly collected from the wild by local farmers or collectors, who often rely only on their experience in identifying the species being collected. Services of specialists like taxonomists are rarely availed for authentication. Thus, it is not uncommon to find admixtures of related/allied species and infrequently also for unrelated genera. Among the reasons attributed for species admixtures are the apparent confusion in vernacular names between indigenous systems of medicine and local dialects, non availability of authentic plant, similarity in morphological features, etc. (Mitra and Kannan, 2007). The possibility of admixtures is particularly high when the species in question co-occurs with morphologically similar species. Frequently, admixtures could also be deliberate due to adulteration (Mitra and Kannan, 2007). The consequences of species admixtures can range from reducing the efficacy of the drug to lowering the trade value (Wieniawski, 2001; Song et al., 2009). Efforts have been made to accurately identify medicinal plants (Jaysinghe et al., 2009). Besides conventional methods including examination of wood anatomy and morpho-taxonomical keys, several-DNA-based methods have been developed to resolve these problems (Sucher and Carles, 2008). With the advent of DNA barcode tools, attempts are being made to use several candidate barcode regions to identify species discussed above.

8. Limitations of DNA barcoding

DNA-based species identification depends on distinguishing intraspecific from interspecific genetic variation. The ranges of these types of variation are unknown and may differ between taxa. It may be difficult to resolve recently diverged species or new species that have arisen through hybridization. There is no universal gene for DNA barcoding, no single gene that is conserved in all domains of life and exhibits enough sequence divergence for species discrimination. The validity of DNA barcoding therefore depends on establishing reference sequences from taxonomically confirmed specimens. This is likely to be a complex process that will involve cooperation among a diverse group of scientists and institutions. Barcode sequences are, in general, short (approx. 500–1000 bp) and this fundamentally limits their utility in resolving deep branches (between orders or phyla) in phylogenies. Some controversy exists over the value of DNA barcoding, largely because of the perception that this new identification method would diminish rather than enhance traditional morphology-based taxonomy, and species determinations based solely on the genetic divergence could result in incorrect species recognition

(Schindel and Miller, 2005). However, we must keep open the possibility that the barcode sequences *per se* and their ever-increasing taxonomic coverage could become an unprecedented resource for taxonomy, systematic and diagnosis (Kress et al., 2005; Chase et al., 2005, 2007), and may be equally useful (Monaghan et al., 2005).

9. Conclusions

In conclusion, the classical way of practice of plant taxonomy for the identification of species lead the discipline many a times to a subject of opinion; the plant DNA barcoding is now transitioning the epitome of species identification.

Acknowledgements

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding the work through the research group project No. RGP-VPP-195. The research was funded in part by the project Excellence in Faculty Support-Research, Centre of Excellence 17586-4/2013/TUD-POL, Hungary. The technical assistance of Sabiha Ali Ajmal is highly appreciated.

References

- Ackerfield, J.R., Wen, J., 2003. Evolution of *Hedera* (the ivy genus, Araliaceae): insights from chloroplast DNA data. *Int. J. Plant Sci.* 164, 593–602.
- Ali, M.A., Choudhary, R.K., 2011. India needs more plant taxonomists. *Nature* 471 (7336), 37–37.
- Ali, M.A., Lee, J., Kim, S.Y., Al-Hemaid, F.M.A., 2012. Molecular phylogenetic study of *Cardamine amaraeformis* Nakai using nuclear and chloroplast DNA markers. *Genet. Mol. Res.* 11 (3), 3086–3090.
- Ali, M.A., Al-Hemaid, F.M.A., Choudhary, R.K., Lee, J., Kim, S.Y., Rub, M.A., 2013. Status of *Reseda pentagyna* Abdallah & A.G. Miller (Resedaceae) inferred from analysis of combined nuclear ribosomal and chloroplast sequence data. *Bangladesh J. Plant Taxon.* 20 (2), 233–238.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25 (17), 3389–3402.
- Alverson, A.J., Wei, X., Rice, D.W., Stern, D.B., Barry, K., Palmer, J.D., 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* 27 (6), 1436–1448.
- Alzohairy, A.M., Gyulai, G., Jansen, R.K., Bahieldin, A., 2013. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. *Plasmid* 69, 1–15.
- Alzohairy, A.M., Yousef, M.A., Edris, S., Kerti, B., Gyulai, G., Bahieldin, A., 2012. Detection of LTR Retrotransposons reactivation induced by *in vitro* environmental stresses in barley (*Hordeum vulgare*) via RT-qPCR. *Life Sci. J.* 9, 5019–5026.
- Andersson, L., Rova, J.H.E., 1999. The rps16 intron and the phylogeny of the Rubioideae (Rubiaceae). *Plant Syst. Evol.* 214, 161–186.
- Angiosperm Phylogeny Group, APGI, 1998. An ordinal classification for the families of flowering plants. *Ann. Miss. Bot. Gard.* 85, 531–553.
- Angiosperm Phylogeny Group, APGII, 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141, 399–436.
- Angiosperm Phylogeny Group, APGIII, 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121.
- Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. *New Biotechnol.* 25 (4), 195–203.
- Armstrong, K.F., Ball, S.L., 2005. DNA barcodes for biosecurity: invasive species identification. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360 (1462), 1813–1823.
- Asahina, H., Shinozaki, J., Masuda, K., Morimitsu, Y., Satake, M., 2010. Identification of medicinal *Dendrobium* species by phylogenetic analyses using *matK* and *rbcL* sequences. *J. Nat. Med.* 64, 133–138.
- Asmussen, C.B., Chase, M.W., 2001. Coding and noncoding plastid DNA in palm systematics. *Am. J. Bot.* 88, 1103–1117.
- Avise, J.C., 1994. *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.
- Baker, W.J., Asmussen, C.B., Barrow, S., Dransfield, J., Hedderson, T.A., 1999. A phylogenetic study of the palm family (Palmae) based on chloroplast DNA sequences from the *trnL-trnF* region. *Plant Syst. Evol.* 219, 111–126.
- Baldwin, B.G., 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol. Phylogenet. Evol.* 1, 3–16.
- Baldwin, B.G., Markos, S., 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S–26S nrDNA: congruence of ETS and ITS trees of Calycadenia (Compositae). *Mol. Phylogenet. Evol.* 10, 449–463.
- Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J., 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Miss. Bot. Gard.* 82, 247–277.
- Baum, D.A., Small, R.L., Wendel, J.F., 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Syst. Biol.* 47, 181–207.
- Besansky, N.J., Severson, D.W., Ferdig, M.T., 2003. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends Parasitol.* 19, 545–546.
- Bhargava, M., Sharma, A., 2013. DNA barcoding in plants: evolution and applications of in silico approaches and resources. *Mol. Phylogenet. Evol.* 67 (3), 631–641.
- Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, Brun, J.J., Taberlet, P., 2012. Tracking earthworm communities from soil DNA. *Mol. Ecol.* 21, 2017–2030.
- Bisset, W.G., 1984. *Herbal Drugs and Phytopharmaceuticals*. CRC Press, London.
- Blaxter, M., 2003. Counting angels with DNA. *Nature* 421, 122–124.
- Blaxter, M., Elsworth, B., Daub, J., 2004. DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Proc. R. Soc. B* 271, 189–192.
- Buckler, E.S., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145, 821–832.
- Bult, C.J., Sweere, J.A., Zimmer, E.A., 1995. Cryptic sequence simplicity, nucleotide composition bias, and molecular coevolution in the large subunit of ribosomal DNA in plants: implications for phylogenetic analyses. *Ann. Miss. Bot. Gard.* 82, 235–246.
- Cai, J.N., Zhou, K.Y., Xu, L.S., Wang, Z.T., Shen, X., Wang, Y.Q., Li, X.B., 2000. Ribosomal DNA ITS sequence analyses of *Cnidium monnieri* from different geographical origin in China. *Acta Pharm. Sin.* 35, 56–59.
- Cai, Z.H., Li, P., Dong, T.T., Tsim, K.W., 1999. Molecular diversity of 5S-rRNA spacer domain in *Fritillaria* species revealed by PCR analysis. *Planta Med.* 65, 360–364.
- Callaway, E., 2012. A bloody boon for conservation. *Nature* 484, 424–425.

- Cameron, S., Rubinoff, D., Will, K., 2006. Who will actually use DNA barcoding and what will it cost? *Syst. Biol.* 55 (5), 844–847.
- Campbell, C.S., Wojciechowski, M.F., Baldwin, B.G., Alice, L.A., Donoghue, M.J., 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier agamic* complex (Rosaceae). *Mol. Biol. Evol.* 14, 81–90.
- Cao, H., Cai, J.N., Liu, Y.P., Wang, Z.T., Xu, L.S., 2001. Correlative analysis between geographical distribution and nucleotide sequence of chloroplast *matK* gene of *Cnidium monnieri* fruit in China. *Chin. Pharm. J.* 36, 373–376.
- Carlsen, T., Aas, A.B., Lindner, D., Vrålstad, T., Schumacher, T., Kanserud, H., 2012. Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol.* 5, 747–749.
- Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., De Mattia, F., 2010. DNA barcoding: a six-question tour to improve users' awareness about the method. *Brief. Bioinform.* 11 (4), 440–453.
- Chandler, G.T., Plunkett, G.M., 2004. Evolution in Apiales: nuclear and chloroplast markers together in (almost) perfect harmony. *Bot. J. Linn. Soc.* 144, 123–147.
- Chase, M.W., Cowan, R.S., Hollingsworth, P.M., Van den Berg, C., Madrinán, S., Petersen, G., Seberg, O., Jørgensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barracough, T.G., Kelly, L., Wilkinson, M., 2007. A proposal for a standardized protocol to barcode all land plants. *Taxon* 56, 295–299.
- Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kes-anakurthi, R.P., Haidar, N., Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long term goals. *Philos. Trans. R. Soc. B* 360, 1889–1895.
- Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.L., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J.D., Manhart, J.R., Systsma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedren, M., Gaut, B.S., Jansen, R.K., Kim, K.J., Wimpee, C.F., Smith, J.F., Furnier, G.R., Strauss, S.H., Xiang, Q.Y., Plunkett, G.M., Soltis, P.S., Swensen, S.M., Williams, S.E., Gadek, P.A., Quinn, C.J., Eguiarte, L.E., Golenberg, E., Learn, G.H., Graham, S.W., Barrett, S.C.H., Dayanandan, S., Albert, V.A., 1993. Phylogenetics of seed plants – an analysis of nucleotide-sequences from the plastid gene *rbcL*. *Ann. Miss. Bot. Gard.* 80, 528–580.
- Che, J., Tang, L., Liu, Y.J., He, W., Chen, F., 2007. Molecular identity of *Crocus sativus* and its misused substitutes by ITS sequence. *China J. Chin. Mater. Med.* 32, 668–671.
- Chen, F., Chan, H.Y., Wong, K.L., Wang, J., Yu, M.T., But, P.P.H., Shaw, P.C., 2008. Authentication of *Saussurea lappa*, an endangered medicinal material, by ITS DNA and 5S *rRNA* sequencing. *Planta Med.* 74, 889–892.
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., Leon, C., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5 (1), e8613.
- Cho, Y., Mower, J.P., Qiu, Y.L., Palmer, J.D., 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci. USA* 101, 17741–17746.
- Choudhary, R.K., Ali, M.A., Lee, J., 2011. Studies on genetic diversity among population of *Persicaria barbata* (L.) Hara from India based on internal transcribed spacer sequences of nuclear ribosomal DNA. *Saudi. J. Biol. Sci.* 18 (2), 123–127.
- Clerc-Blain, J.L., Starr, J.R., Bull, R.D., Saarela, J.M., 2010. A regional approach to plant DNA barcoding provides high species resolution of sedges (*Carex* and *Kobresia*, Cyperaceae) in the Canadian Arctic Archipelago. *Mol. Ecol. Resour.* 10, 69–91.
- Coissac, E., Riaz, T., Puillandre, N., 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.* 21, 1834–1847.
- Collins, R.A., Cruickshank, R.H., 2013. The seven deadly sins of DNA barcoding. *Mol. Ecol. Resour.* 13 (6), 969–975.
- Combet, C., Blanchet, C., Geourjon, C., Deléage, G., 2000. NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25, 147–150.
- Cowen, R.K., Paris, C.B., Srinivasan, A., 2006. Scaling of connectivity in marine populations. *Science* 311 (5760), 522–527.
- Cuenoud, P., Savolainen, V., Chatrou, L.W., Powell, M., Grayer, R.J., Chase, M.W., 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *Am. J. Bot.* 89, 132–144.
- Dasmahapatra, K.K., Mallet, J., 2006. DNA barcodes: recent successes and future prospects. *Heredity* 97 (4), 254–255.
- de Groot, G.A., During, H.J., Maas, J.W., Schneider, H., Vogel, J.C., Erkens, R.H., 2011. Use of *rbcL* and *trnL-F* as a two-locus DNA barcode for identification of NW-European ferns: an ecological perspective. *PLoS ONE* 6 (1), e16371.
- Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., Miaud, C., 2011. Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE* 6 (8), e23389.
- Demesure, B., Sodji, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4, 129–131.
- Dereeper, A.L., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469.
- DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. Biol.* 360, 1905–1916.
- Ding, P., Fang, Q., 2005. Ribosomal DNA-ITS sequence analysis and molecular identification of *Morinda officinalis* and its counterfeit species. *Chin. Trad. Herbal Drugs* 36, 908–911.
- Ding, X., Xu, L., Wang, Z., Zhou, K., Xum, H., Wang, Y., 2002a. Authentication of stems of *Dendrobium officinale* by rDNA ITS region sequences. *Planta Med.* 68, 191–192.
- Ding, X.Y., Wang, Z.T., Xu, H., Xu, L.S., Zhou, K.Y., 2002b. Database establishment of the whole rDNA ITS region of *Dendrobium* species of “Fengdou” and authentication by analysis of their sequences. *Acta Pharm. Sin.* 37, 567–573.
- Ding, X.Y., Wang, Z.T., Xu, L.S., Xu, H., Zhou, K.Y., Shi, G.X., 2002c. Study on sequence difference and SNP phenomenon of rDNA ITS region in F type and H type population of *Dendrobium officinale*. *China J. Chin. Mater. Med.* 27, 85–89.
- Dobzhansky, T., 1973. Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 125–129.
- Dong, L.N., Wortley, A.H., Wang, H., Li, D.Z., Lu, L., 2011. Efficiency of DNA barcodes for species delimitation: a case in *Pterygiella* Oliv. (Orobanchaceae). *J. Syst. Evol.* 49 (3), 189–202.
- Dong, T.T., Ma, X.Q., Clarke, C., Song, Z.H., Ji, Z.N., Lo, C.K., Tsim, K.W., 2003. Phylogeny of *Astragalus* in China: molecular evidence from the DNA sequences of 5S *rRNA* spacer, ITS, and 18S *rRNA*. *J. Agric. Food Chem.* 51, 6709–6714.
- Downie, S.R., Katz-Downie, D.S., 1999. Phylogenetic analysis of chloroplast *rps16* intron sequences reveals relationships within the woody southern African Apiaceae subfamily Apioideae. *Can. J. Bot.* 77, 1120–1135.
- Doyle, J.A., Donoghue, M.J., Zimmer, E.A., 1994. Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Ann. Miss. Bot. Gard.* 81, 419–450.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.

- Du, Z.Y., Qimike, A., Yang, C.F., Chen, J.M., Wang, Q.F., 2011. Testing four barcoding markers for species identification of Potamogetonaceae. *J. Syst. Evol.* 49 (3), 246–251.
- Ebach, M.C., Holdrege, C., 2005. DNA barcoding is no substitute for taxonomy. *Nature* 434, 697.
- Edwards, D., Horn, A., Taylor, D., Savolainen, V., Hawkins, J.A., 2008. DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). *Taxon* 57, 1317–1327.
- Elbadri, G.A., De Ley, P., Waeyenberge, L., Vierstraete, A., Moens, M., Vanfleteren, J., 2002. Intraspecific variation in *Radopholus similis* isolates assessed with restriction fragment length polymorphism and DNA sequencing of the internal transcribed spacer region of the ribosomal RNA cistron. *Int. J. Parasitol.* 32, 199–205.
- Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Malles, J., Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proc. R. Soc.* 274, 2881–2889.
- Epp, L.S., Boessenkool, S., Bellemain, E.P., Haile, J., Esposito, A., Riaz, T., Erséus, C., Gusarov, V.I., Edwards, M.E., Johnsen, A., Senoien, H.K., Hassel, K., Kausarud, H., Yoccoz, N.G., Bråthen, K.A., Willerslev, E., Taberlet, P., Coissac, E., Brochmann, C., 2012. New environmental metabarcodes for analyzing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* 21, 1821–1833.
- Fang, S.G., Wan, Q.H., Fijihara, N., 2002. Formalin removal from archival tissue by critical point drying. *Biotechniques* 33(3), 604, 606, 608–610.
- Farris, J.S., 1970. Methods for computing Wagner trees. *Syst. Zool.* 19, 83–92.
- Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M., Barrett, S.C., 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* 3 (7), e2802.
- Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C., Newmaster, S.G., Hajibabaei, M., Husband, B.C., 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resour.* s1, 130–139.
- Feliner, G.N., Larena, B.G., Aguilar, J.F., 2004. Fine scale geographic structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). *Ann. Bot.* 93, 189–200.
- Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Finet, C., Timme, R.E., Delwiche, C.F., Marlétaz, F., 2010. Multi-gene phylogeny of the green lineage reveals the origin and diversification of land plants. *Curr. Biol.* 21, 2217–2222.
- Fišer, P.Z., Buzan, E.V., 2014. 20 years since the introduction of DNA barcoding: from theory to application. *J. Appl. Genet.* 55 (1), 43–52.
- Floyd, R., Eyualem, A., Papert, A., Blaxter, M.L., 2002. Molecular barcodes for soil nematode identification. *Mol. Ecol.* 11, 839–850.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.
- Freudenstein, J.V., Chase, M.W., 2001. Analysis of mitochondrial *nad1b-c* intron sequences in Orchidaceae: utility and coding of length-change characters. *Syst. Bot.* 26, 643–657.
- Fu, Y.M., Jiang, W.M., Fu, C.X., 2011. Identification of species within *Tetrastigma* (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques. *J. Syst. Evol.* 49 (3), 237–245.
- Fushimi, H., Komatsu, K., Namba, T., Isobe, M., 2000. Genetic heterogeneity of ribosomal RNA gene and *matK* gene in *Panax notoginseng*. *Planta Med.* 66, 659–661.
- Gao, J.P., Wang, Y.H., Qiao, C.F., Chen, D.F., 2003. Ribosomal DNA ITS sequences analysis of the Chinese crude drug fructus schisandrae sphenantherae and fruits of *Schisandra viridis*. *China J. Chin. Mater. Med.* 28, 706–710.
- Gao, Q.B., Zhang, D.J., Duan, Y.Z., Zhang, F.Q., Chen, S.L., 2009. Preliminary studies on ITS sequences of nrDNA from *Rhodiola alsia*. *J. Anhui Agric. Sci.* 37, 49–50.
- Gao, R., Zhang, G., 2013. Potential of DNA barcoding for detecting quarantine fungi. *Phytopathology* 103 (11), 1103–1107.
- Gao, T., Yao, H., Song, J.Y., Zhu, Y.J., Liu, C., Chen, S.L., 2010. Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evol. Biol.* 10, 324.
- Gao, X.X., Xhen, X.Y., Luo, Y.S., 2007. Preliminary study on rDNA ITS sequencing and characteristics of *Citrus medica* L. var. *sarcodactylis* (Noot.) Swingle. *Lishizen Med. Mater. Med. Res.* 5, 162–163.
- Gaut, B.S., Morton, B.R., McCaig, B.C., Clegg, M.T., 1996. Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc. Natl. Acad. Sci. USA* 93, 10274–10279.
- Ge, S., Schaal, B.A., Hong, D.Y., 1997. A reevaluation of the status of *A. lobophylla* based on ITS sequence, with reference to the utility of ITS sequence in *Adenophora*. *Acta Pharm. Sin.* 35, 385–395.
- Ge, X.J., Xiao, K., Li, X.Q., Tang, Y.P., 2008. Analysis of Chishui *Dendrobium nobile* based on rDNA ITS variation. *J. Jiangsu Univ. (Med. Ed.)* 18, 510–512.
- Ge, Y.F., Hang, Y.Y., Xia, B., Wei, Y.L., 2007. Sequencing of *trnL-F* and analysis of interspecific genetic relationship of five medicinal species in *Atractylodes* DC. *J. Plant Resour. Environ.* 16, 12–16.
- Gere, J., Yessoufou, K., Daru, B.H., Mankga, L.T., Maurin, O., van der Bank, M., 2013. Incorporating *trnH-psbA* to the core DNA barcodes improves significantly species discrimination within southern African Combretaceae. *Zookeys* 30 (365), 129–147.
- Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24, 774–786.
- Gregory, T.R., 2005. DNA barcoding does not compete with taxonomy. *Nature* 434, 1067.
- Gu, J., Su, J.X., Lin, R.Z., Li, R.Q., Xiao, P.G., 2011. Testing four proposed barcoding markers for the identification of species within *Ligustrum* L. (Oleaceae). *J. Syst. Evol.* 49 (3), 213–224.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Guo, X., Simmons, M.P., But, P.H., Shaw, P.C., Wang, R.J., 2011. Application of DNA barcodes in *Hedyotis* L. (Spermacoceae, Rubiaceae). *J. Syst. Evol.* 49 (3), 203–212.
- Guo, Y., Tsuruga, A., Yamaguchi, S., Oba, K., Iwai, K., Sekita, S., Mizukami, H., 2006. Sequence analysis of chloroplast *chlB* gene of medicinal *Ephedra* species and its application to authentication of *Ephedra* Herb. *Biol. Pharm. Bull.* 29, 1207–1211.
- Gyulai, G., Humphreys, M., Lagler, R., Szabó, Z., Tóth, Z., Bittsánszky, A., Gyulai, F., Heszky, L., 2006. Seed remains of common millet from the 4th (Mongolia) and 15th (Hungary) centuries: AFLP, SSR and mtDNA sequence recoveries. *Seed Sci. Res.* 16, 179–191.
- Gyulai, G., Szabó, Z., Wichmann, B., Bittsánszky, A., Waters Jr., L., Tóth, Z., Dane, F., 2012. Conservation genetics – Heat Map analysis of nuSSRs of a DNA of archaeological watermelons (Cucurbitaceae, *Citrullus l. lanatus*) compared to current varieties. *Genes Genom. Genom.* 6 (S11), 86–96.
- Hajibabaei, M., 2012. The golden age of metasystematics. *Trends Genet. Evol.* 28 (11), 535–537.
- Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Macke, P.M., Hebert, P.D.N., 2005. Critical factors for assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1959–1967.

- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* 6 (4), e17497.
- Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., Hickey, D.A., 2008. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* 23 (4), 167–172.
- Hajibabaei, M., Smith, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B., Hebert, P.D.N., 2006. A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* 6, 959–964.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41, 95–98.
- Hamby, R.K., Zimmer, E.A., 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis, P.S., Soltis, D.E., Doyle, J.J. (Eds.), *Molecular Systematics of Plants*. Chapman & Hall, New York, pp. 50–91.
- Han, J., Zhu, Y., Chen, X., Liao, B., Yao, H., Song, J., Chen, S., Meng, F., 2013. The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. *Biomed. Res. Int.* 2013, 741476.
- Hansen, A.K., Gilbert, L.E., Simpson, B.B., Downie, S.R., Cervi, A.C., Jansen, R.K., 2006. Phylogenetic relationships and chromosome evolution in tropical *Passiflora* based on cpDNA *trnL/trnT* intergenic spacer sequences and distribution of the chloroplast *rpoC1* intron. *Syst. Bot.* 31, 138–150.
- Hao, M.G., Liu, Z.Q., Wang, J.L., 2004. Application of the sequences of rDNA ITS to identify Chinese drug *Hedyotis diffusa*. *J. Anhui Norm. Univ. (Nat. Sci.)* 27, 188–191.
- Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S., Frazer, K.A., 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* 10 (3), R32.
- Haston, E., Richardson, J.E., Stevens, P.F., Chase, M.W., Harris, D.J., 2009. The Linear Angiosperm Phylogeny Group (LAPG) III: a linear sequence of the families in APG II. *Bot. J. Linn. Soc.* 161, 128–131.
- Hayashi, H., Hosono, N., Kondo, M., Hiraoka, N., Ikeshiro, Y., 1998. Phylogenetic relationship of *Glycyrrhiza* plants based on *rbcL* sequences. *Biol. Pharm. Bull.* 21, 782–783.
- Hayashi, H., Hosono, N., Kondo, M., Hiraoka, N., Ikeshiro, Y., Shibano, M., Kusano, G., Yamamoto, H., Tanaka, T., Inoue, K., 2000. Phylogenetic relationship of six *Glycyrrhiza* species based on *rbcL* sequences and chemical constituents. *Biol. Pharm. Bull.* 23, 602–606.
- Hayashi, H., Miwa, E., Inoue, K., 2005. Phylogenetic relationship of *Glycyrrhiza lepidota*, American licorice, in genus *Glycyrrhiza* based on *rbcL* sequences and chemical constituents. *Biol. Pharm. Bull.* 28, 161–164.
- Hebert, P.D.N., Barrett, R.D.H., 2005. Reply to the comment by L. Prendini on “Identifying spiders through DNA barcodes”. *Can. J. Zool.* 83, 505–506.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA Barcodes. *Proc. Biol. Sci.* 270 (1512), 313–321.
- Hebert, P.D.N., Gregory, T.R., 2005. The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54 (5), 852–859.
- Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. *PLoS Biol.* 2 (10), e312.
- Hershkovitz, M.A., Zimmer, E.A., Hahn, W.J., 1999. Ribosomal DNA sequences and angiosperm systematics. In: Hollingsworth, P.M., Bateman, R.M., Gornall, R.J. (Eds.), *Molecular Systematics and Plant Evolution*. Taylor & Francis, London, UK, pp. 268–326.
- Hillis, D.M., Huelsenbeck, J.P., Swofford, D.L., 1994. Hobgoblin of phylogenetics? *Nature* 369, 363–364.
- Hillis, D.M., Moritz, C., Mable, B.K., 1996. *Molecular Systematics*, second ed. Sinauer Associates, Sunderland, Massachusetts.
- Hilu, K.W., Borsch, T., Muller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T.A.B., Rohwer, J.G., Campbell, C.S., Chatrou, L.W., 2003. Angiosperm phylogeny based on *matK* sequence information. *Am. J. Bot.* 90, 1758–1776.
- Hollingsworth, P.M., Bateman, R.M., Gornall, R.J., 1999. *Molecular Systematics and Plant Evolution*. The Systematics Association Special, vol. 57. Taylor & Francis, London.
- Howard, C., Bremner, P.D., Fowler, M.R., Isodo, B., Scott, N.W., Slater, A., 2009. Molecular identification of *Hypericum perforatum* by PCR amplification of the ITS and 5.8S rDNA region. *Planta Med.* 75, 864–869.
- Huang, Y.J., Chen, J.Y., Su, H.J., Huang, Y.Z., Wan, X.F., 2009. rDNA ITS of *Radix ophiopogonis* and *Radix liriopes* from different regions. *Fujian J. Agric. Sci.* 24, 508–512.
- Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Jansen, R.K., Cai, Z., Raubeson, L.A., Daniel, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* 104, 19369–19374.
- Jansen, R.K., Kaitanis, C., Sasaki, C., Lee, S.-B., Tompkins, J., Alverson, A.J., Daniell, H., 2006. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among Rosids. *BMC Evol. Biol.* 6, 32.
- Janzen, D.J., 2004. Now is the time. *Philos. Trans. R. Soc. Lond. B* 359, 731–732.
- Janzen, D.H., Hajibabaei, M., Burns, J.M., Hallwachs, W., Remigio, E., Hebert, P.D.N., 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philos. Trans. R. Soc. B* 360, 1835–1845.
- Jayasinghe, R.L., Niu, H., Coram, T.E., Kong, S., Kaganovitch, J., Xue, C.C.L., Li, C.G., Pang, E.C.K., 2009. Effectiveness of an innovative prototype subtracted diversity array (SDA) for fingerprinting plant species of medicinal importance. *Planta Med.* 75, 1180–1185.
- Ji, K.P., Li, X.H., Li, Y.D., Zhang, X.L., 2003a. Identification of *Rheum palmatum* L (Dahuang) by method of measuring internal transcribed spacer regions of rRNA gene. *World Sci. Technol. – Modern. Trad. Chin. Med.* 4, 44–47.
- Ji, K.P., Li, Y.D., Zhang, X.L., Li, X.H., 2002. Identification by measuring internal transcribed spacer regions of rRNA gene in *Radix Angelicae Sinensis*. *Chin. Trad. Herbal Drugs* 34, 66–69.
- Ji, K.P., Zhang, X.L., Liu, L.S., Lu, Q.Y., Cheng, C., 2003b. Primary study on measuring the internal transcribed spacer 1 regions of rRNA gene in seeds of *Gentiana dahurica*. *China J. Chin. Mater. Med.* 28, 313–316.
- Jiang, J.H., Meng, N., Cao, X.Y., Zhou, S.B., Dai, C.C., 2005. ITS sequence analysis on medicinal plants of *Euphorbia* L. in Anhui and Jiangsu Provinces. *Chin. Trad. Herbal Drugs* 36, 900–902.
- Jiang, L.Y., Guo, Z.G., Wang, C., Zhao, G.F., 2009. ITS sequence analysis of *Gynostemma pentaphyllum* from different habitats in China. *Chin. Trad. Herbal Drugs* 40, 1123–1127.
- Jiang, R.W., Hon, P.M., Xu, Y.T., Chan, Y.M., Xu, H.X., Shaw, P.C., But, P.P.H., 2006. Isolation and chemotaxonomic significance of tuberostemospirone-type alkaloids from *Stemona tuberosa*. *Phytochemistry* 67, 52–57.
- Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J., Wheeler, T.A., 2014. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Mol. Ecol. Resour.* 14 (2), 221–232.

- Jordan, W.C., Courtney, M.W., Neigel, J.E., 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *Am. J. Bot.* 83, 430–439.
- Kahn, S.D., 2011. On the future of genomic data. *Science* 331, 728–729.
- Kalendar, R., Lee, D., Schulman, A.H., 2009. FastPCR software for PCR primer and probe design and repeat search. *Genes Genom. Genom.* 3, 1–14.
- Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., Schulman, A.H., 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* 97 (12), 6603–6607.
- Keller, A., Förster, F., Müller, T., Dandekar, T., Schultz, J., Wolf, M., 2010. Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. *Biol. Direct.* 5, 4.
- Kelly, L.J., Ameka, G.K., Chase, M.W., 2010. DNA barcoding of African Podostemaceae (river-weeds): a test of proposed barcode regions. *Taxon* 59, 251–260.
- Khatoun, S., Rai, V., Rawat, A.K.S., Mehrotra, S., 2006. Comparative pharmacognostic studies of three *Phyllanthus* species. *J. Ethnopharm.* 104, 79–86.
- Koch, M.A., Haubold, B., Mitchell-Olds, T., 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis* and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498.
- Kojoma, M., Kurihara, K., Yamada, K., Sekita, S., Satake, M., Iida, O., 2002. Genetic identification of cinnamon (*Cinnamomum* spp.) based on the *trnL-trnF* chloroplast DNA. *Planta Med.* 68, 94–96.
- Köljalg, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F., Bahram, M., Bates, S.T., Bruns, T.D., Bengtsson-Palme, J., Callaghan, T.M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G.W., Hartmann, M., Kirk, P.M., Kohout, P., Larsson, E., Lindahl, B.D., Lücking, R., Martin, M.P., Matheny, P.B., Nguyen, N.H., Niskanen, T., Oja, J., Peay, K.G., Peintner, U., Peterson, M., Pöldmaa, K., Saag, L., Saar, I., Schübler, A., Scott, J.A., Sené, C., Smith, M.E., Suija, A., Taylor, D.L., Telleria, M.T., Weiss, M., Larsson, K.H., 2013. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22 (21), 5271–5277.
- Komatsu, K., Zhu, S., Fushimi, H., Qui, T.K., Cai, S., Kadota, S., 2001. Phylogenetic analysis based on 18S rRNA gene and *matK* gene sequences of *Panax vietnamensis* and five related species. *Planta Med.* 67, 461–465.
- Kondo, K., Terabayashi, S., Higuchi, M.I., 1998. Discrimination between *Banxia* and *Tiannanxing* based on *rbcL* sequences. *Nat. Med.* 52, 253–258.
- Kondo, K., Terabayashi, S., Okada, M., 1996. Phylogenetic relationship of medicinally important *Cnidium officinale* and Japanese Apiaceae based on *rbcL* sequences. *J. Plant. Res.* 109, 21–27.
- Kosakovsky, P., Frost, S.D., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679.
- Kress, J.W., Erickson, D.L., 2008. DNA barcodes: genes, genomics, and bioinformatics. *Proc. Natl. Acad. Sci. USA* 105, 2761–2762.
- Kress, J.W., Wurdack, J.K., Zimmer, E.A., Weigt, A.L., Janzen, H.D., 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* 102, 8369–8374.
- Kuzoff, R.K., Sweere, J.A., Soltis, D.E., Soltis, P.S., Zimmer, E.A., 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. *Mol. Biol. Evol.* 15, 251–263.
- Lágler, R., Gyulai, G., Humphreys, M., Szabó, Z., Horváth, L., Bittsánszky, A., Kiss, J., Holly, L., Heszky, L., 2005. Morphological and molecular analysis of common millet (*P. miliaceum*) cultivars compared to an aDNA sample from the 15th century (Hungary). *Euphytica* 146, 77–85.
- Lágler, R., Gyulai, G., Szabó, Z., Tóth, Z., Bittsánszky, A., Horváth, L., Kiss, J., Gyulai, F., Heszky, L., 2006. Molecular diversity of common millet (*P. miliaceum*) compared to archaeological samples excavated from the 4th and 15th centuries. *Hung. Agric. Res.* 1, 14–19.
- Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V., 2008. DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. USA* 105, 2923–2928.
- Laiou, A., Mandolini, L.A., Piredda, R., Bellarosa, R., Simeone, M.C., 2013. DNA barcoding as a complementary tool for conservation and valorisation of forest resources. *Zookeys* 30 (365), 197–213.
- Lambert, D.M., Baker, A., Huynen, L., Haddrath, O., Hebert, P.D.N., Millar, C.D., 2005. Is a large-scale DNA-based inventory of ancient life possible? *J. Hered.* 96, 279–284.
- Lau, D.T., Shaw, P.C., Wang, J., But, P.P.H., 2001. Authentication of medicinal *Dendrobium* species by the internal transcribed spacer of ribosomal DNA. *Planta Med.* 67, 456–460.
- Law, S.K., Simmons, M.P., Techen, N., Khan, I.A., He, M.F., Shaw, P.C., But, P.P.H., 2010. Molecular analyses of the Chinese herb Leigongteng (*Tripterygium wilfordii* Hook.f.). *Phytochemistry* 72, 21–26.
- Ledford, H., 2008. Botanical identities: DNA barcoding for plants comes a step closer. *Nature* 451, 616.
- Lee, E.K., Cibrian-Jaramillo, A., Kolokotronis, S.O., Katari, M.S., Stamatakis, A., Ott, M., Chiu, J.C., Little, D.P., Stevenson, D.W., McCombie, W.R., Martienssen, R.A., Coruzzi, G., Desalle, R., 2011. A functional phylogenomic view of the seed plants. *PLoS Genet.* 7, e1002411.
- Lee, J., Kim, S.Y., Park, S.H., Ali, M.A., 2013. Molecular phylogenetic relationships among members of the family Phytolaccaceae *sensu lato* inferred from internal transcribed spacer sequences of nuclear ribosomal DNA. *Genet. Mol. Res.* 12 (4), 4515–4525.
- Lee, M.S.Y., 2004. The molecularization of taxonomy. *Invert. Syst.* 18, 1–6.
- Leebens-Mack, J., Raubeson, L.A., Cui, L., Kuehl, J., Fourcade, M., Chumley, T., Boore, J.L., Jansen, R.K., dePamphilis, C.W., 2005. Identifying the basal angiosperms in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol. Biol. Evol.* 22, 1948–1963.
- Li, H., Homer, N., 2010. A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinform.* 2 (5), 473–483.
- Li, M., Cao, H., But, P.P.H., Shaw, P.C., 2011a. Identification of herbal medicinal materials using DNA barcodes. *J. Syst. Evol.* 49 (3), 271–283.
- Li, M., Ling, K.H., Lam, H., Shaw, P.C., Cheng, L., Techen, N., Khan, I.A., Chang, Y.S., But, P.P.H., 2010. *Cardiocrinum* seeds as a replacement for *Aristolochia* fruits in treating cough. *J. Ethnopharmacol.* 130, 429–432.
- Li, P., Cai, Z.H., Xing, J.B., 2001. Preliminary attempt to identify geoherbalsism of Flos Lonicerae by sequence divergence of 5S-rRNA gene spacer region. *Chin. Trad. Herbal Drugs* 32, 834–837.
- Li, Y., Gao, L.M., Poudel, R.C., Li, D.Z., Forrest, A., 2011b. High universality of *matK* primers for barcoding gymnosperms. *J. Syst. Evol.* 49 (3), 169–175.
- Lin, S., Zheng, W.W., Wu, J.Z., Zhou, L.J., Song, Y.N., 2007. PCR, clone and sequence analysis of rDNA-ITS of *Nelumbo nucifera* from different geographical origins in China. *China J. Chin. Mater. Med.* 32, 671–675.
- Linder, C.R., Goertzen, L.R., Heuvel, B.V., Francisco-Ortega, J., Jansen, R.K., 2000. The complete external transcribed spacer of 18S–26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. *Mol. Phylogenet. Evol.* 14, 285–303.

- Lipscomb, D., Platnick, N., Wheeler, Q., 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol.* 18 (2), 65–66.
- Little, D.P., 2013. A DNA mini-barcode for land plants. *Mol. Ecol. Resour.* <http://dx.doi.org/10.1111/1755-0998.12194>.
- Liu, J.Q., Chen, Z.D., Liao, Z.X., Lu, A.M., 2001a. A comparison of the ITS sequences of the Tibetan medicine “Zangyinchen”: *Swertia mussotii* and its adulterant species. *Acta Pharm. Sin.* 36, 67–70.
- Liu, J.Q., Chen, Z.D., Lu, A.M., 2001b. Comparison on internal transcribed spacers (ITS) sequences of Tibetan medicine *Saussurea medusa* and its easily confusable species. *Chin. Trad. Herbal Drugs* 32, 443–445.
- Liu, T., Ji, Y.H., 2009. *PsbA-trnH* sequence analysis from chloroplast nonmedicinal plants of *Artemisia*. *Chin. Agric. Sci. Bull.* 25, 46–49.
- Liu, Y.N., Xu, L., Dou, D.Q., 2010. PCR amplification, cloning and sequence analysis of rDNA ITS of *Arctium lappa* from different geographical origins in China. *Chin. Trad. Herbal Drugs* 33, 26–28.
- Liu, Y.P., Cao, H., Han, G.R., Fushimi, H., Komatsu, K., 2002. *MatK* and ITS nucleotide sequencing of crude drug Chuanxiong and phylogenetic relationship between their species from China and Japan. *Acta Pharm. Sin.* 37, 63–68.
- Liu, Z., Zeng, X., Yang, D., Chu, G., Yuan, Z., Chen, S., 2012. Applying DNA barcodes for identification of plant species in the family Araliaceae. *Gene* 499 (1), 76–80.
- Liu, Z.Q., Hao, M.G., 2005. Identification of *Hedyotis diffusa* using rDNAITS sequence. *Shaanxi J. Trad. Chin. Med.* 26, 167–169.
- Lorenz, J.G., Jackson, W.E., Beck, J.C., Hanner, R., 2005. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philos. Trans. R. Soc. B* 360, 1869–1877.
- Luo, K., Chen, S., Chen, K., Song, J., Yao, H., Ma, X., Zhu, Y., Pang, X., Yu, H., Li, X., Liu, Z., 2010. Assessment of candidate plant DNA barcodes using the Rutaceae family. *Sci. China Life Sci.* 53, 701–708.
- Luo, H.B., Zhang, Z., Wu, Z.H., Yun, D.P., 2010. Analysis of nrDNAITS gene sequence of famous-region drug *Codonopsis tangshen* Oliv. *J. Anhui Agric. Sci.* 38, 4427–4428.
- Luo, Y., Yang, Q.E., 2008. Study on ITS sequences of herbal medicine of Chuanwu and Caowu. *Chin. Pharm. J.* 43, 820–823.
- Ma, J., Benetzen, J.L., 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. USA* 101, 12404–12410.
- Ma, X.J., Wang, X.Q., Xiao, P.G., Hong, D.Y., 2000. Comparison of ITS sequences between wild ginseng DNA and garden ginseng DNA. *China J. Chin. Mater. Med.* 25, 206–209.
- Ma, Y.H., Yang, J.A., Jia, W.Z., Ye, G.S., 2004. Sequence analysis of ITS of nuclear ribosomal DNA (nrDNA) of *Eucommia ulmoides* from different geographical origin in China. *J. Northwest Forest. Univ.* 19, 16–19.
- Mallet, J., Willmott, K., 2003. Taxonomy: renaissance or tower of Babel? *Trends Ecol. Evol.* 18 (2), 57–59.
- Manen, J.F., Boulter, M.C., Naciri-Graven, Y., 2002. The complex history of the genus *Ilex* L. (Aquifoliaceae): evidence from the comparison of plastid and nuclear DNA sequences and from fossil data. *Plant Syst. Evol.* 235, 79–98.
- Manen, J.F., Natali, A., 1995. Comparison of the evolution of ribulose-1,5-biphosphate carboxylase (*rbcL*) and *atpB-rbcL* non-coding spacer sequences in a recent plant group, the tribe Rubieae (Rubiaceae). *J. Mol. Evol.* 41, 920–927.
- Mansour, A., Ismail, H.M., Ramadan, M.F., Gyulai, G., 2009a. Comparative genotypic and phenotypic analysis of tomato (*Lycopersicon esculentum*) cultivars grown under two different seasons in Egypt. *Global Science Books, UK. Afr. J. Plant Sci. Biotechnol.* 3 (1), 73–79.
- Mansour, A., Silva, J.A.T., Gyulai, G., 2009b. Assessment of Molecular (Dis)similarity: the role of multiple sequence alignment (MSA) programs in biological research. *Genes Genom. Genom.* 3, 23–30.
- Mao, S.G., Luo, Y.M., Shen, J., Ding, X.Y., 2007. Authentication of *Crocus sativus* L. and its adulterants by rDNAITS sequences and allele-specific PCR. *J. Nanjing Normal Univ. (Nat. Sci. Ed.)* 30, 89–92.
- Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24 (3), 133–141.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgeson, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.L., Jarvie, T.P., Jirage, K.B., Kim, J., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.
- Markmann, M., Tautz, D., 2005. Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. *Philos. Trans. R. Soc. B* 360, 1917–1924.
- Marshall, E., 2005. Will DNA bar codes breathe life into classification? *Science* 307, 1037.
- Martins, L., Hellwig, F.H., 2005. Systematic position of the genera *Serratula* and *Klasea* Centaureinae (Cardueae, Asteraceae) inferred from ETS and ITS sequence data and new combination in *Klasea*. *Taxon* 54 (3), 632–638.
- Mason, B., 2003. Marine surveys sees net gain in number of fish species. *Nature* 425, 889.
- Mathews, S., Donoghue, M.J., 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286, 947–950.
- Metzker, M.L., 2005. Emerging technologies in DNA sequencing. *Genome Res.* 15, 1767–1776.
- Meusnier, I., Singer, G.A.C., Landry, J.F., Hickey, D.A., Hebert, P.D.N., Hajibabaei, M., 2008. A universal DNA minicobarcode for biodiversity analysis. *BMC Genom.* 12 (9), 214.
- Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* 3 (12), e422.
- Mitchell, A., Wen, J., 2005. Phylogeny of *Brassaiopsis* (Araliaceae) in Asia based on nuclear ITS and 5S-NTS DNA sequences. *Syst. Bot.* 30 (4), 872–886.
- Mitra, S.K., Kannan, R., 2007. A Note on unintentional adulterations in ayurvedic herbs. *Ethnobot. Leaflets* 11, 11–15.
- Mizukami, H., 1995. Amplification and sequence of a 5S-rRNA gene spacer region from the crude drug “*Angelica* root”. *Biol. Pharm. Bull.* 18, 1299–1301.
- Mizukami, H., Hao, B.S., Tanaka, T., 1997. Nucleotide sequence of 5S-rDNA intergenic spacer region in *Angelica acutiloba*. *Nat. Med.* 51, 376–378.
- Monaghan, M.T., Balke, M., Gregory, T.R., Vogler, A.P., 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philos. Trans. R. Soc. B* 360, 1925–1933.
- Mondini, L., Noorani, A., Pagnotta, M.A., 2009. Assessing plant genetic diversity by molecular tools. *Diversity* 1, 19–35.
- Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. USA* 104, 19363–19368.
- Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. USA* 107, 4623–4628.

- Moritz, C., Cicero, C., 2004. DNA barcoding: promise and pitfalls. *PLoS Biol.* 2 (10), e354.
- Nair, V.K., Yoganarasimhan, K.R., Murthy, K., Shantha, T.R., 1983. Studies on some South Indian market samples of ayurvedic drugs II. *Ancient Sci. Life* 3, 60–66.
- Nandi, W.I., Chase, M.W., Endress, P.K., 1998. A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets. *Ann. Miss. Bot. Gard.* 85, 137–212.
- Newmaster, S.G., Fazekas, A., Steeves, R., Janovec, J., 2008. Testing candidate plant barcode regions in the Myristicaceae. *Mol. Ecol. Resour.* 8, 480–490.
- Newmaster, S.G., Ragupathy, S., 2009. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Mol. Ecol. Resour.* 9, S172–S180.
- Neyland, R., 2002. A phylogeny inferred from large-subunit (26S) ribosomal DNA sequences suggests the family Dasypogonaceae is closely aligned with Restionaceae allies. *Aust. Syst. Bot.* 15, 749–754.
- Ngan, F., Shaw, P., But, P., Wang, J., 1999. Molecular authentication of *Panax* species. *Phytochemistry* 50, 787–791.
- Niu, X.L., Ji, K.P., Lu, G.Q., 2010. Preliminary studies on identification of *Aquilaria sinensis* (L.) Gilg by the PCR product of rDNA ITS sequencing. *Guangdong Agric. Sci.* 37, 167–169.
- Oxelman, B., Liden, M., Berglund, D., 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Syst. Evol.* 206, 393–410.
- Palmer, J.D., 1992. Mitochondrial DNA in plant systematics. In: Soltis, P.S., Soltis, D.E., Doyle, J.J. (Eds.), *Molecular Systematics of Plants*. Chapman & Hall, New York, pp. 36–49.
- Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L., Song, K., 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc. Natl. Acad. Sci. USA* 97, 6960–6966.
- Palmer, S.A., Smith, O., Allaby, R.G., 2012. The blossoming of plant archaeogenetics. *Ann. Anat.* 194 (1), 146–156.
- Pan, H.X., Hunag, F., Wang, P.X., Zhou, L.J., Cao, L.Y., Liang, R.Y., 2001. Identification of *Amomum villosum*, *Anomum villosum* var. *xanthioides* and *Amomum longiligulare* on ITS-1 sequence. *Chin. Trad. Herbal Drugs* 24, 481–482.
- Pandey, A.K., Ali, M.A., 2006. Molecular markers in plant systematics I: nuclear sequences. In: Kumar, S. (Ed.), *Plant Sciences Research in India: Challenges and Prospects*. Botanical Survey of India, Dehradun, India, pp. 21–34.
- Pandey, A.K., Ali, M.A., 2012. Intraspecific variation in *Panax assamicus* Ban. (Araliaceae) populations based on internal transcribed spacer (ITS1-5.8S-ITS2) sequences of nrDNA. *Ind. J. Biotechnol.* 11, 30–38.
- Pandey, A.K., Ali, M.A., Biate, D.L., Misra, A.K., 2009. Molecular systematics of *Aralia-Panax* complex (Araliaceae) in India based on ITS sequences of nrDNA. *Proc. Natl. Acad. Sci. India* 79, 255–261.
- Pang, X., Shi, L., Song, J., Chen, X., Chen, S., 2013. Use of the potential DNA barcode ITS2 to identify herbal materials. *J. Nat. Med.* 67 (3), 571–575.
- Pang, X.H., Song, J.Y., Zhu, Y.J., Xu, H.X., Huang, L.F., Chen, S.L., 2011. Applying plant DNA barcodes for Rosaceae species identification. *Cladistics* 27, 165–170.
- Pennisi, E., 2007. Taxonomy. Wanted: a barcode for plants. *Science* 318, 190–191.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, Automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* 21, 1864–1877.
- Qin, N.J., Huang, Y., Yang, G., Xu, L.S., Zhou, K.Y., 2003. *RbcL* sequence analysis of *Belamcanda chinensis* and related medicinal plants of *Iris*. *Acta Pharm. Sin.* 38, 147–152.
- Qiu, Y.L., Li, L., Wang, B., Xue, J.Y., Hendry, T.A., Li, R.Q., Brown, J.W., Liu, Y., Hudson, G.T., Chen, Z.D., 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *J. Syst. Evol.* 48, 391–425.
- Rannala, B., Yang, Z., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* 43, 304–311.
- Rao, N.K., 2004. Plant genetic resources: advancing conservation and use through biotechnology. *Afr. J. Biotechnol.* 3 (2), 136–145.
- Rieseberg, L.H., Wood, T.E., Baack, E.J., 2006. The nature of plant species. *Nature* 440, 524–527.
- Ruzicka, J., Lukas, B., Merza, L., Gohler, I., Abel, G., Popp, M., Novak, J., 2009. Identification of *Verbena officinalis* based on ITS sequence analysis and RAPD-derived molecular markers. *Planta Med.* 75, 1271–1276.
- Saarela, J.M., Sokoloff, P.C., Gillespie, L.J., Consaul, L.L., Bull, R.D., 2013. DNA barcoding the Canadian Arctic flora: core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species. *PLoS ONE* 8 (10), e77982.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., De Bruijn, A.Y., Sullivan, S., Qiu, Y.L., 2000. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* gene sequences. *Syst. Biol.* 49, 306–362.
- Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K., Lane, R., 2005. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philos. Trans. R. Soc.* 360, 1805–1811.
- Savolainen, V., Reeves, G., 2004. A plea for DNA banking. *Science* 304, 1445.
- Schindel, D.E., Miller, S.E., 2005. DNA barcoding a useful tool for taxonomists. *Nature* 435, 17.
- Schulman, A.H., 2007. Molecular markers to assess genetic diversity. *Euphytica* 158, 313–321.
- Seberg, O., Humphries, C.J., Knapp, S., Stevenson, D.W., Petersen, G., Scharff, N., Andersen, N.M., 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18, 63–65.
- Seberg, O., Petersen, G., 2009. How many loci does it take to DNA barcode a *Crocus*? *PLoS ONE* 4, e4598.
- Seibel, P., Müller, T., Dandekar, T., Schultz, J., Wolf, M., 2006. 4SALE – a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinform.* 7, 498.
- Seibel, P., Müller, T., Dandekar, T., Wolf, M., 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4 SALE. *BMC Res. Notes* 1, 91.
- Shander, C., Willassen, E., 2005. What can biological barcoding do for marine biology. *Mar. Biol. Resour.* 1, 79–83.
- Shen, Y.J., Yan, P., Zhao, X., Pang, Q.H., Zhao, S.J., 2008. Application of ISSR marker and ITS sequence to investigation of genetic variation of *Aquilaria sinensis*. *J. South China Univ. Technol. (Nat. Sci. Ed.)* 36, 128–132.
- Shen, Y.Y., Chen, X., Murphy, R.W., 2013. Assessing DNA barcoding as a tool for species identification and data quality control. *PLoS ONE* 8 (2), e57125.
- Shi, L.C., Zhang, J., Han, J.P., Song, J.Y., Yao, H., Zhu, Y.J., Li, J.C., Wang, Z.Z., Xiao, W., Lin, Y.L., Xie, C.X., Qian, Z.Z., Chen, S.L., 2011. Testing the potential of proposed DNA barcodes for species identification of Zingiberaceae. *J. Syst. Evol.* 49 (3), 261–266.
- Shi, Z.G., An, W., Zhao, Y.L., 2008. Genetic polymorphism of 18 *Lycium barbarum* resources based on nrDNA ITS sequences. *J. Anhui Agric. Sci.* 49, 10379–10380.
- Shiba, M., Kondo, K., Miki, E., Yamaji, H., Morota, T., Terabayashi, S., Takeda, S., Sasaki, H., Miyamoto, K., Aburada, M.,

2006. Identification of medicinal *Actinophytolobos* based on ITS sequences of nrDNA. *Biol. Pharm. Bull.* 29, 315–320.
- Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* 21, 1794–1805.
- Shokralla, S., Zhou, X., Janzen, D.H., Hallwachs, W., Landry, J.F., Jacobus, L.M., Hajibabaei, M., 2011. Pyrosequencing for mini-barcoding of fresh and old museum specimens. *PLoS ONE* 6 (7), e21252.
- Simmons, M.P., Savolainen, V., Clevinger, C.C., Archer, R.H., Davis, J.I., 2001. Phylogeny of the Celastraceae inferred from 26S nuclear ribosomal DNA, phytochrome B, *rbcL*, *atpB*, and morphology. *Mol. Phylogenet. Evol.* 19, 353–366.
- Smith, M.A., Fisher, B.L., Hebert, P.D.N., 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos. Trans. R. Soc. B* 360, 1825–1834.
- Soltis, D.E., Soltis, P.S., Doyle, J.J., 1998. *Molecular Systematics of Plants II. DNA Sequencing*. Kluwer Academic Publishers, Boston, Massachusetts.
- Soltis, D.E., Soltis, P.S., Nickrent, D.L., Johnson, L.A., Hahn, W.J., Hoot, S.B., Sweere, J.A., Kuzoff, R.K., Kron, K.A., Chase, M.W., Swensen, S.M., Zimmer, E.A., Chaw, S.M., Gillespie, L.J., Kress, W.J., Sytsma, K.J., 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Miss. Bot. Gard.* 84, 1–49.
- Soltis, E.D., Soltis, P.S., 2000. Contributions of plant molecular systematics to studies of molecular evolution. *Plant Mol. Biol.* 42, 45–75.
- Soltis, P.S., Gitzendanner, M.A., 1999. Molecular systematics and the conservation of rare species. *Conserv. Biol.* 13, 471–483.
- Song, J., Yao, H., Li, Y., Li, X., Lin, Y., Liu, C., Han, J., Xie, C., Chen, S., 2009. Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. *J. Ethnopharmacol.* 24, 434–439.
- Spooner, D.M., 2009. DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. *Am. J. Bot.* 96, 1177–1189.
- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Starr, J.R., Naczi, R.F.C., Chouinard, B.N., 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Mol. Ecol. Resour.* 9, S151–S163.
- Stoeckle, M., 2003. Taxonomy, DNA, and the Bar Code of Life. *Bioscience* 53 (9), 796–797.
- Štorchová, H., Olson, M.S., 2007. The architecture of the chloroplast *psbA-trnH* non-coding region in angiosperms. *Plant Syst. Evol.* 268 (1–4), 235–256.
- Su, C., Wong, K.L., But, P.P.H., Su, W.W., Shaw, P.C., 2010. Molecular authentication of the Chinese herb Huajuhong and related medicinal material by DNA sequencing and ISSR marker. *J. Food Drug Anal.* 18, 161–170.
- Sucher, N.J., Carles, M.C., 2008. Genome-based approaches to the authentication of medicinal plants. *Planta Med.* 74, 603–623.
- Sui, X.Y., Hung, Y., Tan, Y., Guo, Y., Long, C.L., 2010. Molecular authentication of the ethnomedicinal plant *Sabia parviflora* and its adulterants by DNA barcoding technique. *Planta Med.* 77 (5), 492–496.
- Sukrong, S., Zhu, S., Ruangrunsi, N., Phadungcharoen, T., Palanuvej, C., Komatsu, K., 2007. Molecular analysis of the genus *Mitragyna* existing in Thailand based on rDNA ITS sequences and its application to identify a narcotic species: *Mitragyna speciosa*. *Biol. Pharm. Bull.* 30, 1284–1288.
- Sun, Y., Fung, K.P., Leung, P.C., Shi, D., Shaw, P.C., 2004. Characterization of medicinal *Epimedium* species by 5S rRNA gene spacer sequencing. *Planta Med.* 70, 287–288.
- Sun, Y., Shaw, P.C., Fung, K.P., 2007. Molecular authentication of *Radix puerariae lobatae* and *Radix puerariae thomsonii* by ITS and 5S rRNA spacer sequencing. *Biol. Pharm. Bull.* 30, 173–175.
- Sunita, G., 1992. *Substitute and Adulterant Plants*. Periodical Experts Book Agency, New Delhi.
- Swofford, D.L., 2002. PAUP: Phylogenetic Analysis using Maximum Parsimony (and other method). Version 4.0b10. Sinauer, Sunderland, Massachusetts.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), *Molecular Systematics*, second ed. Sinauer, Sunderland, Massachusetts, pp. 407–514.
- Szabó, Z., Gyulai, G., Humphreys, M., Horváth, L., Bittsánszky, A., Lágler, R., Heszky, L., 2005. Genetic variation of melon (*C. melo*) compared to an extinct landrace from the Middle Ages (Hungary) I. rDNA, SSR and SNP analysis of 47 cultivars. *Euphytica* 146, 87–94.
- Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012a. Environmental DNA. *Mol. Ecol.* 21, 1789–1793.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012b. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050.
- Taberlet, P., Prud'homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.C., Melodelima, C., Pompanon, F., Coissac, E., 2012c. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.* 21, 1816–1820.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 10, 2731–2739.
- Tao, X.Y., Gui, X.Q., Fu, C.X., Qiu, Y.X., 2008. Analysis of genetic differentiation and phylogenetic relationship between *Changium smyrnioides* and *Chuanminshen violaceum* using molecular markers and ITS sequences. *J. Zhejiang Univ. (Agric. Life Sci.)* 34, 473–481.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.* 18 (2), 70–74.
- Taylor, H.R., Harris, W.E., 2012. An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Mol. Ecol. Resour.* 12, 377–388.
- Techen, N., Parveen, I., Panm, Z., Khan, I.A., 2014. DNA barcoding of medicinal plant material for identification. *Curr. Opin. Biotechnol.* 25, 103–110.
- Teng, Y.F., Wu, X.J., Xu, H., Wang, Z.T., Yu, G.D., Xu, L.S., 2002. A comparison of *matK* sequences between *Herba Dendrobii* (Shihu) and its adulterant species. *J. China Pharm. Univ.* 33, 280–283.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882.
- Valentini, A., Pompanon, F., Taberlet, P., 2009. DNA barcoding for ecologists. *Trends Ecol. Evol.* 24 (2), 110–117.
- Vences, M., Thomas, M., Bonett, R.M., Vieites, D.R., 2005. Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philos. Trans. R. Soc. B* 360, 1859–1868.
- Vongsak, B., Kengtong, S., Vajrodya, S., Sukrong, S., 2008. Sequencing analysis of the medicinal plant *Stemona tuberosa* and five related species existing in Thailand based on *trnH-psbA* chloroplast DNA. *Planta Med.* 74, 1764–1766.
- Wallander, E., Albert, V.A., 2000. Phylogeny and classification of Oleaceae based on *rps16* and *trnL-F* sequence data. *Am. J. Bot.* 87, 1827–1841.

- Wang, D., Bai, Y., Chen, Z.H., 2007. Comparison of ribosomal DNA ITS sequence of *Rhizoma dioscoreae* in different geographical regions. *Lishizen Med. Mater. Med. Res.* 18, 54–55.
- Wang, H., Wang, Q., 2005. Analysis of rDNA ITS sequences of *Radix et Rhizoma Salviae Miltiorrhizae* and plants of *Salvia L.* *Acta Pharm. Sin.* 36, 105–109.
- Wang, Q., Yu, Q.S., Liu, J.Q., 2011. Are nuclear loci ideal for barcoding plants? A case study of genetic delimitation of two sister species using multiple loci and multiple intraspecific individuals. *J. Syst. Evol.* 49 (3), 182–188.
- Wang, S.M., Liang, S.W., Zhou, K.Y., Liu, Z.Q., Feng, W.S., Wu, M.X., 2004. Ribosomal rDNA ITS sequence analysis of root of *Achyranthes bidentata*. *Acta Pharm. Sin.* 35, 559–562.
- Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B* 360, 1847–1857.
- Wen, J., Pandey, A.K., 2005. Initiating DNA molecular systematic studies in a developing country. In: Pandey, A.K., Wen, J., Dogra, J.V.V. (Eds.), *Plant Taxonomy: Advances and Relevance*. CBS Publishers & Distributors, New Delhi, India, pp. 31–43.
- White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), *PCR Protocols: A Guide to Method and Amplifications*. Academic Press, San Diego, California, pp. 315–322.
- Wieniawski, W., 2001. Risk assessment as an element of drug control. *WHO Drug Inform.* 15, 7–11.
- Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54 (5), 844–851.
- Will, K.W., Rubinoff, D., 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20, 47–55.
- Wilson, E.O., 2003. The encyclopaedia of life. *Trends Ecol. Evol.* 18, 77–80.
- Wolf, M., Ruderisch, B., Dandekar, T., Schultz, J., Müller, T., 2008. ProfDistS: (profile-) distance based phylogeny on sequence-structure alignments. *Bioinformatics* 24, 2401–2402.
- Wyman, S.K., Boore, J.L., Jansen, R.K., 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255.
- Xiang, X.G., Zhang, J.B., Lu, A.M., Li, R.Q., 2011. Molecular identification of species in Juglandaceae: a tiered method. *J. Syst. Evol.* 49 (3), 252–260.
- Xie, H., Huo, K.K., Chao, Z., Pan, S.L., 2009. Identification of crude drugs from Chinese medicinal plants of the genus *Bupleurum* using ribosomal DNA ITS sequences. *Planta Med.* 75, 89–93.
- Xie, H., Zhao, Z.L., Huo, K.K., Wu, B.Y., Pan, S.L., 2006. ITS sequence of 9 *Bupleurum* species and its application in identification of Chaihu (*Radix Bupleuri*). *J. Southern Med. Univ.* 26, 1460–1463.
- Xu, H., Li, X.B., Wang, Z.T., Ding, X.Y., Xu, L.S., Zhou, K.Y., 2001. RDNAITS sequencing of *Herba Dendrobii* (hünagcao). *Acta Pharm. Sin.* 36, 777–783.
- Xu, H., Wang, Z., Ding, X., Zhou, K., Xu, L., 2006. Differentiation of *Dendrobium* species used as “Huangcao Shihu” by rDNA ITS sequence analysis. *Planta Med.* 72, 89–92.
- Xue, C.Y., Li, D.Z., 2011. Use of DNA barcode sensu lato to identify traditional Tibetan medicinal plant *Gentianopsis paludosa* (Gentianaceae). *J. Syst. Evol.* 49 (3), 267–270.
- Yan, Hai-Fei, Hao, G., Hu, C.M., Ge, X.J., 2011. DNA barcoding in closely related species: a case study of *Primula L.* sect. *Proliferae Pax* (Primulaceae) in China. *J. Syst. Evol.* 49 (3), 225–236.
- Yan, P., Pang, Q.H., Jiao, X.W., Zhao, X., Shen, Y.J., Zhao, S.J., 2008. Genetic variation and identification of cultivated *Fallopia multiflora* and its wild relatives by using chloroplast *matK* and 18S rRNA gene sequences. *Planta Med.* 74, 1504–1509.
- Yang, D.Y., Fushimi, H., Cai, S.Q., Komatsu, K., 2004. Molecular analysis of *Rheum* species used as Rhei Rhizoma based on the chloroplast *matK* gene sequence and its application for identification. *Biol. Pharm. Bull.* 27, 375–383.
- Yang, Y., Zhai, Y., Liu, T., Zhang, F., Ji, Y., 2010. Detection of *Valeriana jatamansi* as an adulterant of medicinal *Paris* by length variation of chloroplast *psbA-trnH* region. *Planta Med.* 77, 81–97.
- Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Yang, Z., Rannala, B., 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol. Biol. Evol.* 14, 717–724.
- Yang, Z.Y., Chao, Z., Huo, K.K., Xie, H., Tian, Z.P., Pan, S.L., 2007. ITS sequence analysis used for molecular identification of the *Bupleurum* species from northwestern China. *Hytomedicine* 14, 416–423.
- Yao, H., Song, J.Y., Ma, X.Y., Liu, C., Li, Y., Xu, H.X., Han, J.P., Duan, L.S., Chen, S.L., 2009. Identification of *Dendrobium* species by a candidate DNA barcode sequence: the chloroplast *psbA-trnH* intergenic region. *Planta Med.* 75, 667–669.
- Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., Chen, S., 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE* 5 (10), e13102.
- Yoccoz, N.G., 2012. The future of environmental DNA in ecology. *Mol. Ecol.* 21, 2031–2038.
- Yu, J., Xue, J.H., Zhou, S.L., 2011. New universal *matK* primers for DNA barcoding angiosperms. *J. Syst. Evol.* 49 (3), 176–181.
- Yu, M.T., Wong, K.L., Zong, Y.Y., Shaw, P.C., Che, C.T., 2008. Identification of *Swertia mussotii* and its adulterant *Swertia* species by 5S rRNA gene spacer. *China J. Chin. Mater. Med.* 33, 502–504.
- Yu, Y.B., Qin, M.J., Liang, Z.T., Yu, G.D., Tan, N.H., 2003. Ribosomal DNA ITS sequence comparisons of *Pseudostellaria heterophylla* from different geographical regions. *J. Plant Res. Environ.* 12, 1–5.
- Zeng, M., Ma, Y.J., Zheng, S.Q., Xu, J.F., Di, X.H., 2003. Studies on ribosomal DNA sequence analyses of *Radix puerariae* and its sibling species. *Chin. Pharm. J.* 38, 173–175.
- Zhang, H.Y., Shi, X.G., 2007. Analysis of rDNA ITS sequences in root tuber of *Polygonum multiflorum* from various habitats. *Chin. Trad. Herbal Drugs* 38, 911–914.
- Zhang, J.M., Wang, J.X., Xia, T., Zhou, S.L., 2009. DNA barcoding: species delimitation in tree peonies. *Sci. China C Life Sci.* 52, 568–578.
- Zhang, N., Yan, B., Xu, X.H., Xu, L.C., 2010a. Sequence analysis of rDNA-ITS of Baishouwu from different species. *China J. Chin. Mater. Med.* 35, 1537–1540.
- Zhang, W., Han, Y.L., Zhu, J.H., 2010b. Analysis of rDNA ITS sequences of plants of *Aconitum L.* and its sibling species. *J. Biol.* 27, 50–52.
- Zhang, X.L., Ji, K.P., Li, Y.D., Chen, C., Li, X.H., Liu, L.H., 2003. Studies on establishing rRNA gene map of *Angelica sinensis* and *Rheum palmatum* from Gansu by DNA sequencing. *J. Chin. Med. Mater.* 26, 481–483.
- Zhang, Y., Zhang, J.C., Huang, M.H., Yang, M.S., Cao, H., 2006. Detection of genetic homogeneity of *Panax notoginseng* cultivars by sequencing nuclear 18S rRNA and plastid *matK* genes. *Planta Med.* 72, 860–862.
- Zhao, G.P., Niizeki, M., Ishigawa, L., Qian, S.Q., Zhang, Q., Ao, J.N., 2006. ITS Sequence analysis of Chinese and Japanese medicinal plants of *Angelica L.* *Acta Pharm. Sin.* 37, 1072–1076.
- Zhao, H.G., Zhou, J.J., Cao, S.S., Zheng, Y.H., Shan, Y., Xia, B., 2009. Analysis of interspecific relationship among *Stellaria media* and its related species based on ITS and *trnL-F* sequence differences. *J. Plant Resour. Environ.* 18, 1–5.
- Zhao, K.J., Dong, T.T., Cui, X.M., Tu, P.F., Tsim, K.W., 2003. Genetic distinction of radix adenophorae from its adulterants by the DNA sequence of 5S-rRNA spacer domains. *Am. J. Chin. Med.* 31, 919–926.

- Zhao, Z.L., Leng, C.H., Wang, Z.T., 2007. Identification of *Dryopteris crassirhizoma* and the adulterant species based on cpDNA *rbcL* and translated amino acid sequences. *Planta Med.* 73, 1230–1233.
- Zhao, Z.L., Zhou, K.Y., Dong, H., Xu, L.S., 2001. Characters of nrDNA ITS region sequences of fruits of *Alpinia galanga* and their adulterants. *Planta Med.* 67, 381–383.
- Zhao, Z.L., Zhou, K.Y., Wang, Z.T., Qin, M.J., Dong, H., 2000. Determination of nrDNA ITS sequences of Caodoukou (*Alpinia hainanensis* K. Schum.) and Yizhi (*A. oxyphylla* Miq.). *J. Plant Resour. Environ.* 9, 38–40.
- Zhou, L., Wang, P.X., Huang, F., Cao, L.Y., Liang, R.Y., 2002. ITS sequence analysis of *Amomum villosum*. *Chin. Trad. Herbal Drugs* 33, 72–75.
- Zhu, X.Y., Chase, M., Qiu, Y.L., Kong, H.Z., Dilcher, D., Li, J.H., Chen, Z.D., 2007. Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evol. Biol.* 7, 217.
- Zhu, S., Fushimi, H., Cai, S.Q., 2003. A new variety of genus *Panax* from southern Yunnan, China and its nucleotide sequences of 18S ribosomal RNA gene and *matK* gene. *J. Jpn. Bot.* 78, 86–94.
- Zhu, Y., Qin, M.J., Hang, Y.Y., Wang, L.Q., 2007. Authentication of *Pseudostellaria heterophylla* and its counterfeit species by analysis of rDNAITS sequences. *Chin. J. Nat. Med.* 5, 211–215.
- Zimmer, E.A., Martin, S.L., Beverly, S.M., Kan, W., Wilson, A.C., 1980. Rapid duplication and loss of genes coding for the chains of hemoglobin. *Proc. Natl. Acad. Sci. USA* 77, 2158–2162.
- Zuo, Y., Chen, Z., Kondo, K., Funamoto, T., Wen, J., Zhou, S., 2011. DNA barcoding of *Panax* species. *Planta Med.* 72 (2), 182–187.
- Zwickl, D., 2006. Genetic algorithm approaches for the phylogenetic analysis of large Biological sequence datasets under the maximum likelihood criterion (Thesis). Univ. Texas at Austin.