
Use of Organelle Markers to Study Genetic Diversity in Soybean

Lidia Skuza, Ewa Filip and Izabela Szućko

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/52028>

1. Introduction

Soybean is the most important crop provider of proteins and oil used in animal nutrition and for human consumption. Plant breeders continue to release improved cultivars with enhanced yield, disease resistance, and quality traits. It is also the most planted genetically modified crop. The narrow genetic base of current soybean cultivars may lack sufficient allelic diversity to counteract vulnerability to shifts in environmental variables. An investigation of genetic relatedness at a broad level may provide important information about the historical relationship among different genotypes. Such types of study are possible thanks to different markers application, based on variation of organelle DNA (mtDNA or cpDNA).

2. Mitochondrial genome

2.1. Genomes as markers

Typically, all sufficiently variable DNA regions can be used in genetic studies of populations and in interspecific studies. Because of in seed plants chloroplasts and mitochondria are mainly inherited uniparentally, organelle genomes are often used because they carry more information than nuclear markers, which are inherited biparentally. The main benefit is that there is only one allele per cell and per organism, and, consequently, no recombination between two alleles can occur. With different dispersal distances, genomes inherited biparentally, maternally and paternally, also reveal significant differences in their genetic variability among populations. In particular, maternally inherited markers show diversity within a population much better [1].

In gymnosperms the situation is somewhat different. Here, chloroplasts are inherited mainly paternally and are therefore transmitted through pollen and seeds, whereas mitochondria are largely inherited maternally and are therefore transmitted only by seeds [2]. Since pollen is distributed at far greater distances than seeds [3], mitochondrial markers show a greater population diversity than chloroplast markers and therefore serve as important tools in conducting genetic studies of gymnosperms [4]. Mitochondrial markers are also sometimes used in conjunction with cpDNA markers [5].

Mitochondrial regions used in interspecific studies of plants, mainly gymnosperms, include, for example, introns of the NADH dehydrogenase gene *nad1* [4, 5, 6], the *nad7* intron 1 [7], the *nad5* intron 4 [3] and an internally transcribed spacer (ITS) of mitochondrial ribosomal DNA [8, 9].

In addition to the aforementioned organelle markers, microsatellite markers [10, 11] and simple sequence repeats (SSR) are often used in population biology, and sometimes also in phylogeographic studies. Microsatellites are much less common in plants than in animals [12]. However, they are present in both the nuclear genome and the organelle genome. Microsatellites may reveal a high variability, which may be useful in genetic studies of populations, whereas other sequences or methods such as fingerprinting do not detect mutations sufficiently [9,10,13]. Inherited only uniparentally, organelle markers have a certain quality in phylogeographic analyses. Since they are haploid, the effective population size should be reduced after the analysis using these markers as compared to those in which nuclear markers are used [1, 14]. Smaller effective populations sizes should bring about faster turnover rates for newly evolving genotypes, resulting in a clearer picture of past migration history than those obtained using nuclear markers [15-17].

Initially, it was mainly in phylogeographic studies of animal species that mitochondrial markers were used [18]. These studies have provided some interesting data on the beginnings and the evolutionary history of human population [19]. In contrast to studies of animals, using mitochondrial markers in studies of plants, especially angiosperms, is limited [20]. Presently, cpDNA markers are most commonly used in phylogeographic studies of angiosperms, whereas mitochondrial markers are prevalent in studies of gymnosperms.

2.2. Plant mitochondrial DNA

Mitochondrial genomes of higher plants (208-2000 kbp) are much larger than those of vertebrates (16-17 kbp) or fungi (25-80 kbp) [21, 22]. In addition, there are clear differences in size and organization of mitochondrial genomes between different species of plants. Intramolecular recombination in mitochondria leads to complex reorganizations of genomes, and, in consequence, to alternating arrangement of genes, even in individual plants, and the occurrence of duplications and deletions are common [23]. In addition, the nucleotide substitution rate in plant mitochondria is rather low [24], causing only minor differences within certain loci between individuals or even species. Extensively characterized circular animal mitochondrial genomes are highly conservative within a given species; they do not contain introns and have a very limited number of intergenic sequences [25]. Plant mitochondrial DNA (mtDNA) contains introns in multiple genes and several additional genes undergoing

expression when compared to animal mitochondria, but most of the additional sequences in plants are not expressed and they do not seem to be essentials [26]. The completely sequenced mitochondrial genomes are available for several higher plants, including *Arabidopsis thaliana* [27] or *Marchantia polymorpha* [28].

Restriction maps of nearly all plant mitochondrial genomes provide for the occurrence of the master circle with circular subgenomic molecules that arise after recombination among large direct repeats (> 1 kbp) [21, 29-36], which are present in most mitochondrial genomes of higher plants. However, such molecules, whose sizes can be predicted, are very rare or very difficult to observe. It can be explained by the fact that plant mitochondrial genomes are circularly permuted as in the phage T4 [37, 38]. Oldenburg and Bendich reported that mostly linear molecules in *Marchantia* mtDNA are circularly permuted with random ends [39]. It shows that plant mtDNA replication occurs similarly to the mechanism of recombination in the T4 [38].

Many reports that have appeared in recent years indicate that mitochondrial genome of yeasts and of higher plants exist mainly as linear and branched DNA molecules with variable size which is much smaller than the predicted size of the genomes [39-44]. Using pulsed field gel electrophoresis (PFGE) of in-gel lysed mitochondria from different species revealed that only about 6-12% of the molecules are circular [41, 44]. The observed branched molecules are very similar to the molecules seen in yeast in the intermediate stages of recombination of mtDNA [45] or the phage T4 DNA replication [37, 38].

In all but one known case (*Brassica hirta*) [46], plant mitochondrial genomes contain repeat recombinations. These sequences, ranging in length from several hundred to several thousand nucleotides (nt) exist at two different loci in the master circle, yet in four mtDNA sequence configurations [47]. These four configurations correspond to the reciprocal exchange of sequences 5' and 3' surrounding the repeat in the master circle, which suggests that the repeat mediates homologous recombination. Depending on the number and orientation of repeats, the master circle is a more or less complex set of subgenomic molecules [48].

Maternally inherited mutations, which are associated with mitochondria in higher plants, most often occur as a result of intra- and intergenic recombination. This happens in most cases of cytoplasmic male sterility (cms) [41, 49-51], in *chm*-induced mutation in *Arabidopsis* [52] and in non-chromosomal stripe mutations in maize [53]. In this way, it is assumed that the recombination activity explains the complexity of the variations detected in the mitochondrial genomes of higher plants.

2.3. Mitochondrial genome of soybean

The size of soybean mtDNA has been estimated to be approximately 400 kb [54-56]. Spherical molecules have also been observed by electron microscopy [55, 57].

Repeated sequences 9, 23 and 299 bp have been characterized in soybean mitochondria [58, 59]. Also, numerous reorganizations of genome sequences have been characterized among different cultivars of soybean. It has been demonstrated that they occur through homologous recombination produced by these repeat sequences [58, 60, 61], or through short elements that are part of 4.9kb PstI fragment of soybean mtDNA [62]. The 299 bp repeat

sequence has been found in several copies of mtDNA of soybean and in several other higher plants, suggesting that this repeated sequence may represent a hot spot for recombination of mtDNA in many plant species [59, 62]. Previous results suggested that active homologous recombinations of mtDNA are present in at least some species of plants. Recently (2007) amitochondrial-targeted homolog of the *Escherichia coli* *recA* gene in *A. thaliana* has been identified [63]. However, the data on recombination activity in plant mitochondria is still missing. The first data on such an activity in soybean was obtained in 2006 [64]. This discovery is supported by an analysis of mtDNA of soybean using electron microscopy and 2D-electrophoresis. The results suggest that only a small portion of mtDNA molecules undergoes recombination at any given time. Therefore the question is whether this recombination is essential to the functioning of mitochondria and to plant growth.

The repeated sequences of the *atp6*, *atp9* and *coxII* genes have been also characterized, but their recombination activity has not been analysed [65].

The first data for the restriction map of soybean mtDNA were obtained from the analysis of loci of the *atp4* gene [48]. In the vicinity of this gene two repeated sequences that show characteristics of recombination repeats have been found [47, 48]. Active recombination repeats were also identified in circular molecules smaller than 400 kb [55, 66]. These observations suggest that soybean mtDNA has multipartite structure that is similar to other plant mitochondrial genomes containing recombination repeats.

In the mitochondrial genome of cultivar Williams 82, recombinantly active repeats 1 kb and 2 kb have been described [48]. In a different repeat of 10 kb, surrounding both 1 kb and 2 kb repeats, two breakpoints have been identified. This recombination of smaller and larger repeats probably leads to the complex structure of genomes.

The analysis of restriction fragment length polymorphism (RFLP) of mtDNA seems to be a useful method in studying phylogenetic relationships within species.

Grabau et al. (1992) analyzed the genomes of 138 soybean cultivars [60]. Using 2.3 kb *HindIII* mtDNA probe from Williams 82 soybean cultivar revealed restriction fragment length polymorphisms (RFLPs), which allowed for the division of many soybean cultivars into four cytoplasmic groups: Bedford, Arksoy, Lincoln and soja-forage.

Subsequent analyses showed variations within, and adjacent to, the 4.8 kb repeats. Bedford cytoplasm turned out to be the only one that contains copies of the repeat in four different genomic environments, which indicates its recombination activity [61]. Lincoln and Arksoy cytoplasm contain two copies of the repeat and a unique fragment that appear to result from rare recombination events outside, but near, the repeat. In contrast, forage-soja cytoplasm contains no complete repeat, but it contains a unique truncated version of the repeat [61]. Sequence analysis revealed that truncating is caused by the recombination with a repeat of 9 bp CCCCTCCCC. The structural reorganization that occurred in the region around 4.8 kb repeat may provide a way to analyze the relationships between species and evolution within the soybean subgenus.

In order to determine the sources of cytoplasmic variability, Hanlon and Grabau (1995) studied the old cultivars of soybeans with the same 2.3-kb *HindIII* fragment and with a

mtDNA fragment containing the *atp6* gene [62]. They showed that mtDNA RFLP analysis with these probes is useful for the classification of mitochondrial genomes of soybean. Grabau and Davies (1992) made a general classification of wild soybean using the 2.3-kb *HindIII* as a probe [68].

Mt type	Probe	<i>coxI</i>			<i>coxII</i>			<i>atp6</i>		Reference
		Enzyme	<i>HindIII</i>	<i>BamHI</i>	<i>EcoRI</i>	<i>HindIII</i>	<i>BamHI</i>	<i>EcoRI</i>	<i>BamHI</i>	
Ic					1,6	5,8		5,0		[69]
Id					1,6	5,8		5,0;6,0;12,0		[69]
Ie					1,6	5,8		5,0;12,0		[69]
Ik					1,6	5,8		5,0;5,4;5,8		[69]
Ilg					1,3	7,0		1,0;2,6		[69]
IIIa					1,2	8,5		2,4;5,0		[69]
IIIb					1,2	8,5		2,9;5,0		[69]
IIIc					1,2	8,5		5,0;6,0;12,0		[69]
IVa					3,5	8,1		2,4;5,0		[69]
IVb					3,5	8,1		2,9;5,0		[69]
IVc					3,5	8,1		5,0		[69]
IVd					3,5	8,1		2,4;3,5;5,0		[69]
IVe					3,5	8,1		2,6;2,9		[69]
IVf					3,5	8,1		5,2;12,0		[69]
Va					5,8	8,1		2,4;5,0		[69]
V'j					5,8	15,0		5,0;6,0		[69]
VIg					1,7	5,8		1,0;2,6		[69]
VIIg					8,5	15,0		1,0;2,6		[69]
mtI					1,6	5,8				[69]
mtII					1,3	7,0				[69]
mtIII					1,2	8,1				[69]
mtIV					3,5					[69]
mtV					5,8					[69]
mt-a								2,4;5,0		[87]
mt-b								2,9;5,0		[87]
mt-c								5,0		[87]
mt-d								5,0;6,0;12,0		[87]
mt-e								5,0;12,0		[87]

Mt type	Probe	<i>coxI</i>			<i>coxII</i>			<i>atp6</i>		Reference
		Enzyme	<i>HindIII</i>	<i>BamHI</i>	<i>EcoRI</i>	<i>HindIII</i>	<i>BamHI</i>	<i>EcoRI</i>	<i>BamHI</i>	
mt-f								2,4; 3,5; 5,0		[87]
mt-g								1,0; 2,6		[87]
mt-h								2,6; 2,9		[87]
mt-m								2,9		[87]
mt-n								12,0		[87]
Ic		5,6	0,8; 2,5; 5,0	10,5	1,6	5,8	1,9	5,0	8,2; 12,0	[58]
Id		5,6	0,8; 2,5; 5,0	10,5	1,6	5,8	1,9	5,0; 6,0; 12,0	2,8; 6,0; 12,0	[58]
Ie		5,6	0,8; 2,5; 5,0	10,5	1,6	5,8	1,9	5,0; 12,0	2,8; 6,0; 12,0	[58]
Ik		5,6	0,8; 2,5; 5,0	10,5	1,6	5,8	1,9	5,0; 5,4; 5,8	2,8; 6,0; 12,0	[58]
Ilg		8,5	0,8; 2,5; 5,0	9,0	1,3	7,0	4,8	1,0; 2,6	2,8; 3,0; 9,5	[58]
IIIb		5,6	0,8; 2,5; 5,0	10,5	1,2	8,5	6,2; 6,5	2,9; 5,0	6,0; 8,2; 12,0	[58]
IIId		5,6	0,8; 2,5; 5,0	10,5	1,2	8,5	6,2; 6,5	5,0; 6,0; 12,0	3,2; 6,2; 12,0	[58]
Iva		5,6	0,8; 2,5; 5,0	10,5	3,5	8,1	5,0	2,4; 5,0	3,0; 6,0; 12,0	[58]
IVb		5,6	0,8; 2,5; 5,0	10,5	3,5	5,8	5,0	2,9; 5,0	6,0; 8,2; 12,0	[58]
IVc		5,6	0,8; 2,5; 5,0	10,5	3,5	5,8	5,0	5,0	8,2; 12,0	[58]
IVf		5,6	0,8; 2,5; 5,0	10,5	3,5	5,8	5,0	2,4; 3,5; 5,0	3,2; 6,2; 12,0	[58]
IVh		5,6	0,8; 2,5; 5,0	10,5	3,5	5,8	5,0	2,6; 2,9	3,2; 6,2; 12,0	[58]
IVi		5,6	0,8; 2,5; 5,0	10,5	3,5	5,8	5,0	5,2; 12,0	3,2; 6,2; 12,0	[58]
Va		5,6	0,8; 2,5; 5,0	10,5	5,8	5,8	12,0	2,4; 5,0	3,0; 6,0; 12,0	[58]
Vb		5,6	0,8; 2,5; 5,0	10,5	5,8	5,8	12,0	2,9; 5,0	6,0; 8,2; 12,0	[58]
Vc		5,6	0,8; 2,5; 5,0	10,5	5,8	5,8	12,0	5,0	8,2; 12,0	[58]
V'j		5,6	0,8; 2,5; 5,0	10,5	5,8	15,0	1,6	5,0; 6,0	2,8; 6,0; 12,0	[58]

Mt type	Probe	<i>coxI</i>		<i>coxII</i>			<i>atp6</i>		Reference	
		Enzyme	<i>HindIII</i>	<i>BamHI</i>	<i>EcoRI</i>	<i>HindIII</i>	<i>BamHI</i>	<i>EcoRI</i>		<i>BamHI</i>
Vlg		5,6; 8,5	0,8; 2,5; 5,0; 9,0; 10,5	1,7	5,8	4,5	1,0; 2,6	2,8; 3,0; 4,3; 9,5; 12,0	[58]	
VIIg		8,5	0,8; 5,0; 5,2	9,0	8,5	15,0	1,6	1,0; 2,6	2,8; 3,0; 9,5	[58]
VIIIc		5,6	0,8; 2,5; 5,0	10,5	8,5; 10,0	11,0; 15,0	1,6	5,0	8,2; 12,0	[58]
Combined chloroplast and mitochondrial genome type										
cpl +mtIIIb					1,2	8,5		2,9; 5,0		[89]
cpl +mtIVb					3,5	8,1		2,9; 5,0		[89]
cpl+mtIVc					3,5	8,1		5,0		[89]
cplI +mtIVb					3,5	8,1		2,9; 5,0		[89]
cplI +mtIVc					3,5	8,1		5,0		[89]
cplII+mtIe					1,6	5,8		5,0; 12,0		[89]
cplII +mtIVa					3,5	8,1		2,4; 5,0		[89]
cplII +mtVIIIc					8,5; 10,0	11,0; 15,0		5,0		[89]

Table 1. Classification of mitochondrial genome types based on RFLPs using *coxI*, *coxII* and *atp6* as probes. Sizes of hybridization signals (kb) are shown.

In their research Tozuka et al. (1998) used two fragments of mtDNA as probes: the 0.7-kb *HindIII-NcoI* fragment containing the *coxII* (the gene encoding the mitochondrial cytochrome oxidase subunit II) of wild soybean and the 0.66-kb *StyI* fragment containing the *atp6* (the gene encoding the mitochondrial ATPase subunit 6) from *Oenothera* [69, 70] (Table 1).

Based on the RFLPs detected in gel-blot analysis with the *coxII* and *atp6* probes, the harvested plants were divided into 18 groups. Five mtDNA types were described in 94% of the surveyed plants. The geographical distribution of mtDNA types revealed that in many regions soybean growing wild in Japan consisted of a mixture of plants with different types of mtDNA, sometimes even within a single location. Some of these mtDNA types have shown marked geographic clines among the regions. In addition, some wild soybeans had mtDNA types that were identical to those described in cultivated soybeans. These results suggest that mtDNA analysis could resolve maternal origin among of the genus *Glycine* subgenus *Soja* [69].

Kanazawa et al. (1998) gathered 1097 *G. soja* plants from all over Japan and analyzed their RFLP of mitochondrial DNA (mtDNA) using five probes (*coxI*, *coxII*, *atp6*, *atp9*, *atp1=atpA*) [58] (Table 1). 20 different types of mitochondrial genomes labeled as combinations of types I to VII and types from a to k were identified and characterized in this study. Nearly all the mtDNA types described for soybean cultivars also occurred in wild soybean.

The mitochondrial *atpA* gene was also analysed [48]. It was shown that in soybean this gene has a sequence in 90-97% identical with mitochondrial genes of other plants [71-81]. Sequence similarity is limited to the *atpA* coding region. An intriguing feature of the *atpA* open reading frame of soybean is a 642 nt overlap in the putative translation termination site onto an unidentified open reading frame of the *orf214*. The ends of the open reading frame contain four tandems of UGA codon that covers four tandems of AUG codon that initiates an unidentified *orf214* frame. The *atpA-orf214* region was found in soybean mtDNA in multiple sequence contexts. This can be attributed to the presence of two recombination repeats.

The open reading frame shares 79% of nucleotide identity with the *orf214* and is located in the same *atpA* locus position as in common bean *orf209* [82]. Since such organization is a repeat of overlapping the *atpB* and *atpE* reading frames in several chloroplast genes [83, 84], the probability that the *orf214* codes a different ATPase subunit cannot be evaluated because small ATPase subunits are poorly conserved [85].

So for a total of 26 mtDNA haplotypes of wild soybeans have been identified based on RFLP with probes from two mitochondrial genes: *cox2* and *atp6* [69, 86] (Table 1). The three most common haplotypes (Id, IVa and Va) are present in 43 populations. The distribution of mtDNA haplotypes varies among populations [87]. Recently Shimamoto (2001) analyzed the genetic polymorphisms of mitochondrial genes subgenus *Soja* originating from China and Japan [88] (Table 1). As a result of these studies, 6 types of mitochondrial genomes were distinguished.

3. Chloroplast genome

As the result of the extensive research conducted in the past two decades, cpDNA analysis brought about fundamental changes to the systematics of plants. The chloroplast genome is ideal for phylogenetic analyses of plants for several reasons. First, it occurs abundantly in plant cells and is taxonomically ubiquitous. And since it is well researched, it can be easily tested in the laboratory conditions and analyzed in comparative programs. Moreover, it often contains marker structural features cladistically useful, and, above all, it exhibits moderate or low rate of nucleotide substitution [89]. In regard to the mitochondrial genome, and also to cpDNA, researchers use in their studies two distinct phylogenetic approaches [90], namely taxonomic checking of specific traits features of molecular cpDNA and sequencing of specific genes or regions.

3.1. Chloroplast genome of soybean

In estimating the phylogeny of plants belonging to *Glycine*, particular attention was paid to unusual and specific features of cpDNA. In the course of many studies on the variability of

chloroplast genome, a breakthrough came in 1993, with a study on assessing phylogeny of seed plants. The study used a huge database of the nucleotide sequences of the *rbcl* gene [91], encoding the ribulose-1,5-bisphosphate carboxylase, large subunit. The accumulation of a number of comparative data on this chloroplast gene made it a frequent object of research. This is due to the fact that this gene's locus is large (> 1400 bp), and provides many phylogenetically informative traits. The rate of the *rbcl* evolution proved to be appropriate for assessing issues related to phylogeny of plants, especially on the medium and high taxonomic levels. Over the years other sequences from other species as well as many other genes with another chloroplast *atpB* gene coding H⁺ -ATPase subunits [92-95]. The *atpB-rbcL* sequence reaches different lengths in *Glycine* as well as in other seed plants. The study by Chiang (1998) shows that the size of the *atpB-rbcL* space in the studied species ranges from 524 bp to 1000 bp [5], where in the non-coding region the occurrence of deletions and insertions, as well as a number of nucleotide substitutions is a common phenomenon, which can also be observed in *Glycine*. In *Glycine max*, its chloroplast genome differs from the core set chloroplast DNA genes because of the presence of a single, large inversion of approximately 51 kb, in the area between the *rbcl* gene and the *rps16* intron [96]. This inversion is also present in other legumes: the mutation was reported in *Lotus* and *Medicago* [96]. In addition, the non-coding *atpB-rbcL* region is rich in AT, due to which most non-coding regions rich in these base pairs show a small number of functions [97, 98]. Therefore, this predisposes them for faster evolution, and hence for use in molecular systematics.

The summary phylogeny was based on sequence of several cpDNA genes from hundreds of spermatophytes including *Glycine* (Table 2). These genes can be divided into three classes. The genes encoding the photosynthetic apparatus structure form the first class. The second class includes the rRNA genes and genes encoding the chloroplast genetic apparatus. The last class consists of an average of about 30 tRNA encoding genes [99], although their number can vary from 20 to 40 [100, 101].

Genes	Products
Genes for the photosynthesis system	
<i>rbcl</i>	Ribulose -1,5- bisphosphate carboxylase, large subunit
<i>psaA, B</i>	Photosystem I, P700 apoproteins A1, A2
<i>psaC</i>	9kDa protein
<i>psaA</i>	Photosystem II, D1 protein
<i>psaB</i>	47kDa chlorophyll a-binding protein
<i>psaC</i>	43 kDa chlorophyll a-binding protein
<i>psaD</i>	D2 protein
<i>psaE</i>	Cytochrome b559 (8kDa protein)
<i>psaF</i>	Cytochrome b559 (4kDa protein)
<i>psaH</i>	10 kDa phosphoprotein

Genes	Products
Genes for the photosynthesis system	
<i>psal, J, K, L, M, N</i>	-J, -K, -L, -M, -N-proteins
<i>atpA, B, E</i>	H ⁺ -ATPase, CF ₁ subunits α , β , ϵ
<i>atp F, H, I</i>	CF ₀ subunits I, III, IV
<i>petB, D</i>	Cytochrome b ₆ / <i>f</i> complex, subunit b ₆ , IV
<i>nadA- K</i>	NADH Dehydrogenase, subunits ND 1, NDI 1
Genes for the genetic system	
<i>16S rRNA</i>	16S rRNA
<i>23S rRNA</i>	23S rRNA
<i>trnA -UGC</i>	Alanine tRNA (UGC)
<i>trnG- UCC</i>	Glycine tRNA (UCC)
<i>trnH- GUG</i>	Histidine tRNA (GUG)
<i>trnI- GAU</i>	Isoleucine tRNA (GAU)
<i>trnK- UUU</i>	Lysine tRNA (UUU)
<i>trnL- UAA</i>	Leucine tRNA (UAA)
<i>rps2, 7, 12, 16</i>	30S: ribosomal proteins CS2, CS7, CS12, CS16
<i>rp12, 20, 32</i>	50S: ribosomal proteins CL2, CL 20, CL32
<i>rpoA, B, C1, C2</i>	RNA polymerase, subunits α , β , β' , β''
<i>matK</i>	Maturase -like protein
<i>sprA</i>	Small plastid RNA
Others	
<i>clpP</i>	ATP-dependent protease, proteolytic subunit
<i>irf168 (ycf3)</i>	Intron- containing Reading frame (168 codons)

Table 2. Chloroplast genes for the photosynthesis system, for the genetic and others.

The complete size of the *Glycine max* chloroplast genome is 152,218 bp. It contains 25,574 bp of inverted repeats (IRa and IRb), which are separated by a unique small single copy (SSC) region (17,895 bp) [98]. In addition, this genome consists of a large single region (LSC) of unique sequences with 83,175 bp. The IR extends from the *rps19* gene up to the *ycf1*. The *Glycine* chloroplast genome contains 111 unique genes and 19 duplicate copies in the IR, amounting to a total of 130 genes. The cpDNA analysis has showed the presence of 30 different tRNAs in it and 7 of them are repeated within the IR regions. The genes are composed in 60% of encoding regions (52% are protein coding genes and 8% are RNA genes), and in 40% of non-coding regions, including both intergenic spacers and introns. The total content of GC and AT pairs in the *Glycine* chloroplast genes is 34% and 66% respectively. Distinctly

higher percentage of AT pairs (70%) was observed in non-coding regions than in coding regions (62% AT) [98].

In comparison with other eukaryotic genomes, cpDNA is highly concentrated, for example, only 32% of the rice genome is non-coding. In *Glycine max* it is slightly more – 40%. Most of the non-coding DNA is found in very short fragments that separate functional genes. Some studies have shown complex patterns of mutational changes in the non-coding regions. Some of the best known regions in the chloroplast genome is the farther region of the *rbcL* gene in many legumes. This non-coding sequence is flanked by the *rbcL* and *psaI* (the gene encoding the polypeptide I of photosystem I).

3.2. Extent of IR in Glycine

Analysis of the IR (inverted repeats) regions in *Glycine max* has shown that they are separated by a large region and a small region of a unique sequence. In cpDNA repeated sequences are usually located asymmetrically, which results in the formation of long and short regions of a unique sequence [102]. The IR in *Glycine* is a region with 25,574 bp containing 19 genes. At the IR/LSC junction, at the ends of the 5' IR, there is the repeated *rps19* gene (68 bp), and at the junction of the IR/SSC and 5' ends the duplicated *ycf1* gene (478 bp) is located. In the course of study it was shown that comparing cpDNA IR region in *Medicago*, *Lotus*, *Glycine* and *Arabidopsis* indicates that there are changes within the IR in the two legumes. *Glycine* and *Lotus* have 478 bp and 514 bp of the *ycf1* duplicated, whereas *Arabidopsis* has 1,027 bp duplicated in the IR. This contraction of the IR in these legumes accounts for the smaller size of their IR and larger size of the SSC. In addition, contraction of the IR boundary in legumes, IRa has been lost in *Medicago*. This loss has resulted in *ndhF* (usually located in the SSC) being adjacent to *trnH* (usually the first gene in the LSC at the LSC/IRa junction). Loss of one copy of the IR in some legumes provides support for monophyly of six tribes [103-106]. Wolfe (1988) identified duplicated sequences of portions of two genes, 40 bp of *psbA* and 64 bp of *rbcL*, in the region of the IR deletion between *trnH* and *ndhF* in *Pisum sativum* and these duplications were later identified in broad bean (*Vicia faba*) [104,107]. According to many researchers, the IR region is considered the most conserved part of the chloroplast genome, and thus, it is responsible for stabilizing the plastid DNA molecules [108, 109]. Thus the loss of IR can be phylogenetically informative at the local level, as well as misleading at the global phylogeny level, because the IR loss likely occurred independently in more than one group of plants. Coniferous and some legumes (*Pisum sativum*, *Vicia faba*, *Medicago sativa*), for example, contain only one IR. Perhaps the lack of repeat sequences in these plants is associated with an increased incidence of rearrangement of chloroplast genomes [109].

Introns or intergenic sequences in legume chloroplast DNA have become extremely important tools in phylogenetic analyses aimed at systematizing of this species [110, 111]. Moreover, their microstructural changes occur with great frequency in the regions of cpDNA. The body of existing research suggests that mutations in the non-coding regions and relatively fast evolution of the organelle genome encoding regions can serve as valuable markers for the separation species in their evolutionary origin [110, 111]. The systematics of plants gen-

erally considers chloroplast indels to be phylogenetic markers, because of their low prevalence in comparison with nucleotide substitutions [5].

3.3. CpDNA markers

There are many methods of generating molecular markers that rely on site-specific amplification of a selected DNA fragment using polymerase chain reaction (PCR) and its further processing (restriction analysis, sequencing). Initially the research on the plant genome (mostly phylogenetic studies) used non-coding and coding sequences of chloroplast DNA. With time, the genes or DNA segments located in the nuclear DNA, mitochondrial (mtDNA) and chloroplast (cpDNA) found a prominent place among plant DNA markers. Fully automated DNA sequencing made it possible to subject ever-newer regions of plant DNA to comparative sequencing.

One of the most frequently sequenced cpDNA fragments in plant phylogeny of spermatophytes is the *rbcL* gene encoding a large ribulose biphosphate carboxylase subunit (RUBISCO), whose length in most plants is 1,428, 1,431 and 1,434 bp, and insertions and deletions within it are extremely rare [94]. For many years this gene has been the subject of many comprehensive phylogenetic analyses of subgenus *Glycine* [112-114]. The *rbcL* is most commonly used in the analyses at the family and genus levels, but there also exists research at the lower levels, cultivars and wild soybean [98, 115, 116]. A marker with very similar characteristics to those of the *rbcL* (the rate of evolution, the length of 1497 bp) is a gene encoding the ATP synthase β subunit – the *atpB* [94]. The *matK* gene sequence, encoding maturase involved in splicing of the type II introns, and whose length is 1,550 bp is characterized by a rapid rate of evolution that allows to use it in research at the species and genus levels [117, 118]. Frequent mutations in this gene make it unsuitable for studies at higher taxonomic levels. Other popular cpDNA sequences used in phylogenetic studies of legumes include the *ndhF* (the gene encoding the NADH protein, which is a dehyd98rogenase subunit), 16S rDNA, the non-coding *atpB-rbcL* region [94], or the *trnL* (UAA) intron and mediator between the *trnL* (UAA) exon and the *trnF* (GAA) gene [96, 117- 119].

It should be noted that the rate of evolution for a specific DNA region to be used as a marker can vary significantly not only among systematic groups, but also within these groups [98]. Moreover, each DNA fragment within the same group has a different rate of evolution, such as the *ndhF* cpDNA sequence in the *Solanaceae* family, which provides about 1.5 times more information in terms of parsimony than the *rbcL* [90]. Therefore each gene or any other DNA fragment used as a genetic marker has a typical range of "taxonomic" or phylogenetic applications, which can vary significantly within a taxon. For this reason, the *rbcL* sequence has been widely used in *Glycine* for many years at the species and genus levels [104, 117, 118].

3.4. The genetic diversity of soybeans

The importance of genetic variations in facilitating plant breeding and/or conservation strategies has long been recognized [121]. Molecular markers are useful tools for assaying genetic variation and provide an efficient means to link phenotypic and genotypic variation [122].

In recent years, the progress made in the development of DNA based marker systems has advanced our understanding of genetic resources. These molecular markers are classified as: (i) hybridization based markers i.e. restriction fragment length polymorphisms (RFLPs), (ii) PCR-based markers i.e. random amplification of polymorphic DNAs (RAPDs), amplified fragmentlength polymorphisms (AFLPs), inter simple sequence repeats (ISSRs) and micro-satellites or simple sequence repeats (SSRs), and (iii) sequence based markers i.e. single nucleotide polymorphisms (SNPs) [121, 123]. Majority of these molecular markers have been developed either from genomic DNA library (e.g. RFLPs or SSRs) or from random PCR amplification of genomic DNA (e.g. RAPDs) or both (e.g. AFLPs) [123]. Availability of an array of molecular marker techniques and their modifications led to comparative studies among them in many crops including soybean, wheat and barley [124-126]. Among all these, SSR markers have gained considerable importance in plant genetics and breeding owing to many desirable attributes including hypervariability, multiallelic nature, codominant inheritance, reproducibility, relative abundance, extensive genome coverage (including organellar genomes), chromosome specific location, amenability to automation and high throughput genotyping [127]. In contrast, RAPD assays are not sufficiently reproducible whereas RFLPs are not readily adaptable to high throughput sampling. AFLP is complicated as individual bands are often composed of multiple fragments mainly in large genome templates [123]. The general features of DNA markers are presented in Table 3.

	Molecular markers			
	EST-SSRs	SSRs	RFLPs	RAPDs/AFLPs/ISSRs
Need for sequence data	Essential	Essential	Not required	Not required
Level of polymorphism	Low	High	Low	Low-moderate
Dominance	Co-dominant	Co-dominant	Co-dominant	Co-dominant
Interspecific transferability	High	Low-moderate	Moderate-high	Low-moderate
Utility in Marker assisted selection	High	High	Moderate	Low-moderate
Cost and labour involved in generation	Low	High	High	Low-moderate

Table 3. Important features of different types of molecular markers.

The genetic diversity of wild and cultivated soybeans has been studied by various techniques including isozymes [128], RFLP [87], SSR markers [124], and cytoplasmic DNA markers [87, 128, 129]. Based on haplotype analysis of chloroplast DNA, cultivated soybean appears to have multiple origins from different wild soybean populations [129, 130].

Using PCR-RFLP method soybean chloroplast DNAs were classified into three main haplotype groups (I, II and III) [113, 130, 131]. Type I is mainly found in the species of cultivated soybean (*Glycine max*), while types II and III are often found in both the cultivated and wild

forms of soybean (*Glycine soja*). Type III is by far the most dominant in the wild soybean species [113]. In *Glycine*, these types are widely used in evaluating cpDNA variability and in determining phylogenetic relationships between different types of cpDNA using different marker systems. According to Chen and Hebert (1999) [133] analysis of cpDNA sequence is not sufficient for when the analysis of population genetics, and so cpDNA polymorphism assessment methods must be constantly complemented with methods such as single-strand conformation polymorphism (SSCP) [134], or dideoxy fingerprinting (ddF) [135], and directed termination and polymerase chain reaction (DT-PCR). However, some researchers point out that there are many disadvantages of these methods, mainly because of their high cost and large amount of work necessary for obtaining the results. In their view, a single change in the regions of *Glycine* chloroplast DNA at the species and genus levels should be located on a local-specific markers, for example, non-coding regions, using PCR and sequencing.

Analyses of non-coding regions of cpDNA have been employed to elucidate phylogenetic relationship of different taxa [90]. Compared with coding regions, non-coding regions may provide more informative characters in phylogenetic studies at the species level because of their high variability due to the lack of functional constraints. Non-coding regions of cpDNA have been assayed either by direct sequencing [136-141], or by restriction-site analysis of PCR products (PCR-RFLP) [142-146]. In Small's opinion (1998) non-coding regions, which include introns and intergenic sequences, often show greater variability at nucleotides than at the encoding regions, which makes the non-coding regions good phylogenetic markers [139]. Mutations in the form of insertions and deletions are accumulated in noncoding regions at the same rate as nucleotide substitutions, and such kinds of mutations significantly accelerate changes in these regions. In many cases, insertions or deletions are related to short repeat sequences. Therefore, many researchers continually focus on the analysis of non-coding regions. Using RFLP method, Close et al. (1989) found six cpDNA haplotypes and described them in types, ranging from group I to VI, including cultivated and wild soybeans [147]. In the course of their research they found that groups I and II diverge from groups III to VI, thus dividing subgenus *Soja* into two main groups. They presented a hypothesis that group II can be distinguished from group III by two independent mutations. Similar groups of haplotypes in legumes were also obtained by Shimamoto et al. (1992) [128] and Kanazawa et al. (1998) [148], using a combination of *EcoRI* and *ClaI* RFLPs. In their classification, Kazanawa et al. (1998) relied on sequential analysis and found that differences in the three types described by Shimamoto et al. (1992) resulted from two single-base substitutions: one in the non-coding region, between the *rps11* and *rpl36*, and the other in the 3' part of the coding region of the *rps3*. Based on the existing reports, Xu et al. (2000) sequenced nine non-coding regions of cpDNA for seven cultivars and 12 wild forms of soybean (*Glycine max*, *Glycine soja*, *Glycine tabacina*, *Glycine tomentella*, *Glycine microphylla*, *Glycine clandestina*) in order to verify earlier classification of *Glycine* [113]. In the course of their studies, they located eleven single-base changes (substitutions and deletions) in the collected 3849 database. They located five mutations in the distinguished haplotypes I and II, and seven mutations in type III. In addition, haplotypes I and II were identical and clearly different from the taxons in type III. This research has not yielded significant results, because different types of cpDNA could not originate monophyletically, but it contributed to finding a common ances-

tor in the course of evolution of *Glycine*. A neighbor joining tree resulting from the sequence data revealed that the subgenus *Soja* connected with *Glycine microphylla*, which formed a distinct clad from *Clycine clandestine* and the tetraploid cytotypes of *Glycine tabacina* and *Glycine tomentella*. Several informative length mutations of 54 to 202 bases, due to insertions or deletions, were also detected among the species of the genus *Glycine*.

3.5. Non-coding regions of the chloroplast genome as site-specific markers in *Glycine*

In the chloroplast genomes of legumes, including soybean, there are many non-coding regions, which are characterized by a faster rate of evolution when compared to the coding regions. As mentioned earlier some of the chloroplast genes have introns, yet their structure differs from those occurring in the nuclear genes, since in the case of cpDNA introns have a tendency to adopt secondary structure, which affects the model in which cpDNA introns evolve and it is enforced by the secondary structure. This restriction in changes caused by mutations affects the functional requirements related to the formation of introns [98, 108]. As there are no adequate studies on the evolution of introns, it can be assumed that their evolution is similar to that of the protein-encoding genes. The loss of introns in the course of the evolution of chloroplast DNA is an interesting process. It has been discovered that *O. sativa* has 3 introns less in cpDNA than *M. polymorpha* and *N. tabacum*. The loss of an intron in the *rpl2* gene was researched in 340 species representing 109 families of angiosperms including *Glycine* [149]. When trying to determine the taxonomic position, the absence of this intron in a given gene shows that it was lost at least six times in the evolution of angiosperms. In *Glycine* 23 introns have been identified while in *Arabidopsis thaliana* there are 26 introns, mostly located in the same genes and in the same locations within those genes [98, 102].

Non-coding regions in chloroplast DNA have become a major source for phylogenetic studies within the species *Glycine* and in many other seed plants. Earlier, the most popular phylogeny sequences included encoding regions, such as the *rbcL* gene sequences that were designed to determine the phylogenetic relationships between species in major taxonomic groups [113, 136-141]. According to Taberlet et al. (1991) [119] the potential ability of non-coding regions of cpDNA was reserved for species located in the lower taxonomic levels while the non-coding regions, which include introns and intergenic sequences, often show greater variability at nucleotides than is evident in the coding regions, which predisposes them to be used in population studies involving *Glycine*, and others [139,142].

As the result, many studies on phylogenetic utility of non-coding regions have been published [110]. For example [150]: *trnH-psbA*, *trnS-tang*; [148]: *rps11-rpl36*, *rpl16-rps3*, [113]: *trnT-trnL*.

In cpDNA analysis of many plants, very conservative regions flanking areas with high variability are used. The more conservative regions, the higher the chance for the primers designed in the PCR reaction, which will be able to join the broader taxonomic group [96, 113]. The region occurring between the *trnT* (UGU) and the *trnF* (GAA) genes is a large single copy wick is suitable because of the conservativeness of the *trn* genes and several hundred base pairs of noncoding regions. The intergenic space between the *trnT* (UGU) and the *trnL* (UAA) 5' exon ranged from 298 bp to about 700 bp in the species studied by [119]. In the plant genomes completely sequenced by Sugiute, the length of this region is different and

amounts to 770 bp in rice and 710 bp in tobacco. In *Marchantia polymorpha* it is 188 bp [151]. This region is located between the tRNA genes, just as the non-coding sequence located between the *trnL* (UAA) 3' exon and the *trnF* (GAA). Due to its catalytic properties and its secondary structure, the *trnL* (UAA) intron, which belongs to type I introns, is less variable and therefore of better utility for evolutionary studies at higher taxonomic levels [113]. Moreover, depending on the species, they show high frequency of insertions or deletions, which makes them potentially useful as genetic markers.

Region	Primer sequence (5 - 3)*	Annealing temperature	Reference
<i>trnH-psbA</i>	f: TGATCCACTTGGCTACATCCGCC r: GCTAACCTTGGTATGGAAGT	60°C	[99] (tobacco) [150] (soybean)
<i>trnS-trnG</i>	f: GATTAGCAATCCGCCGCTTT r: TTACCACTAAACTATAACCCGC	60°C	[99] (tobacco) [99] (tobacco)
<i>trnT-trnL</i>	f: GGATTGAACCGATGACCAT r: TTAAGTCCGTAGCGTCTACC	60°C	[113] (soybean) [99] (tobacco)
<i>trnL-trnF</i>	f: TCGTGAGGGTTCAAGTCC r: AGATTTGAACTGGTGACACG	56°C	[99] (tobacco) [99] (tobacco)
<i>atpB-rbcL</i>	f: GAAGTAGTAGGATTGATTCTC r: CAACACTTGCTTAGTCTCTG	58°C	[99] (tobacco) [99] (tobacco)
<i>psbB-psbH</i>	f: AGATGTTTTTGCTGGTATTGA r: TTCAACAGTTTGTGTAGCCA	56°C	[99] (tobacco) [99] (tobacco)
<i>rps11-rpl36</i>	f: GTATGGATATATCCATTTTCGTG r: TGAATAACTTACCCATGAATC	50°C	[148] (soybean) [148] (soybean)
<i>rpl16-rps3</i>	f: ACTGAACAGGCGGGTACA r: ATCCGAAGCGATGCGTTG	50°C	[148] (soybean) [148] (soybean)
<i>ndhD-ndhE</i>	f: GAAAATTAAGGAACCCGCAA r: TCAACTCGTATCAACCAATC	56°C	[99] (tobacco) [99] (tobacco)

*f, forward primer; r, reverse primer

Table 4. Primers used for amplification of nine non-coding regions of soybean cpDNA.

In most studied species, the *trnL* (UAA) intron ranges in size from 254 - 767 bp. Its smaller fragment – the P6 loop – reaches a length of 10 - 143 bp. It is commonly applied in DNA barcoding. Its main limitation lies in its low homologousness with the species from the Gene Bank, which amounts to 67.3%, while the homologousness of the P6 loop is 19.5%. However, it also has some advantages: conservative primers projected form and trouble-free amplification process. Amplification of the P6 loop can be performed even in a very degraded DNA. The intron is well known and its sequences are used to determine phylogenetic relationships between closely related species or to identify a plant species [152]. The first universal primers for this region were designed more than 20 years ago [119]. However, it does not

belong to the most variable non-coding regions in chloroplast DNA [108]. The *trnL* (UAA) intron is the only one belonging to group I introns in chloroplast DNA, which means that its secondary structure is highly conservative, with a possibility of changes in its conservative [113] and variability in regions [99, 153]. Consequently, comparing the diversity of the *trnL* intron sequences allows to obtain new primers that contain conservative regions and amplify short sections contained between them [152].

Thus, in angiosperms, using non-coding regions in research at lower levels of the genome is a routine practice [108]. A large number of non-coding regions of cpDNA has been located in angiosperms, some of which are highly variable, whereas others show relatively small variability [108]. In studying the chloroplast genome, many researchers looked for universal primers that would allow amplification of many non-coding regions of cpDNA (Table 4) [111, 113, 148, 150].

4. Conclusion

In phylogenetic and population studies of *Glycine*, genetic information contained not only in cpDNA but also in mtDNA are often analysed. Organelle DNA can be used to find species-specific molecular markers. Molecular markers are an important tool to systematize the species because their use allows for detecting the differences in the genes directly. The selection of appropriate sequences, which depends on the taxonomic level at which reconstruction of the origin is carried out is very important. The initial selection concerns non-conservative sequences, which are subject to fast evolution, because the more related the specimen are, the more changeable the region should be. The relatively slow rate of evolution of certain sequences may exclude statistically significant analyses within families or species, while the study of relationship between species, which phylogenetically are very distant, using more slowly evolving sequences can be very useful. Non-coding sequences show a faster rate of evolution than the coding sequences. These regions accumulate a greater number of insertion/deletion or substitution than the non-coding regions, and therefore may be more suitable for research at inter-or intra-genus levels.

Author details

Lidia Skuza*, Ewa Filip and Izabela Szućko

*Address all correspondence to: skuza@univ.szczecin.pl

Cell Biology Department, Faculty of Biology, University of Szczecin, Poland

References

- [1] Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., & Vendramin, G. G. (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, 14, 689-701.
- [2] Wagner, D. B. (1992). Nuclear, chloroplast, and mitochondrial DNA polymorphisms as biochemical markers in population genetic analyses of forest trees. *New Forests*, 6, 373-390.
- [3] Liepelt, S., Bialozyt, R., & Ziegenhagen, B. (2002). Wind-dispersed pollen mediates postglacial gene flow among refugia. USA. *Proceedings of the National Academy of Sciences*, 99, 14590-14594.
- [4] Johansen, A. D., & Latta, R. G. (2003). Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. *Molecular Ecology*, 12, 293-298.
- [5] Chiang, Y. C., Hung, K. H., Schaal, B. A., Ge, X. J., Hsu, T. W., & Chiang, T. Y. (2006). Contrasting phylogeographical patterns between mainland and island taxa of the *Pinus luchuensis* complex. *Molecular Ecology*, 15, 765-779.
- [6] Jaramillo-Correa, J. P., Beaulieu, J., & Bousquet, J. (2004). Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology*, 13, 2735-2747.
- [7] Godbout, J., Jaramillo, J. P., Beaulieu, J., & Bousquet, J. (2005). A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molecular Ecology*, 14, 3497-3512.
- [8] Huang, S., Chiang, Y. C., Schaal, B. A., Chou, C. H., & Chiang, T. Y. (2001). Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. *Molecular Ecology*, 10, 2669-2681.
- [9] Pleines, T., Jakob, S. S., & Blattner, D. B. (2009). Application of non-coding DNA regions in intraspecific analyses. *Plant Systematic and Evolution*, 282, 281-294.
- [10] Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, 17, 6463-6471.
- [11] Tautz, D., Trick, M., & Dover, G. A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature*, 322, 652-656.
- [12] Lagercrantz, U., Ellegren, H., & Andersson, L. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research*, 21, 1111-1115.
- [13] Powell, W., Morgante, M., Mc Devitt, R., Vendramin, G. G., & Rafalski, J. A. (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to

- the population genetics of pines. USA. *Proceedings of the National Academy of Sciences*, 92, 7759-77763.
- [14] Birky, C. W., Fuerst, P., & Maruyama, T. (1989). Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, 121, 613-627.
- [15] Rendell, S., & Ennos, R. A. (2002). Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. *Molecular Ecology*, 11, 69-78.
- [16] Hudson, R. R., & Coyne, J. A. (2002). Mathematical consequences of the genealogical species concept. *Evolution*, 56, 1557-1565.
- [17] Kadereit, J. W., Arafah, R., Somogyi, G., & Westberg, E. (2005). Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. *Taxon*, 54, 861-876.
- [18] Avise, J. C. (2000). *Phylogeography: the history and formation of species*, Cambridge, Harvard University Press.
- [19] Richards, M. B., Macaulay, V. A., Bandelt, H. J., & Sykes, B. C. (1998). Phylogeography of mitochondrial DNA in western Europe. *Annals of Human Genetics*, 62, 241-260.
- [20] Tomaru, N., Takahashi, M., Tsumura, Y., Takahashi, M., & Ohba, K. (1998). Intraspecific variation and phylogeographic patterns of *Fagus crenata* (Fagaceae) mitochondrial DNA. *American Journal of Botany*, 85, 629-636.
- [21] Mackenzie, S., & Mc Intosh, L. (1999). Higher plant mitochondria. *Plant Cell*, 11, 571-585.
- [22] Schuster, W., & Brennicke, A. (1994). The plant mitochondrial genome: physical structure, information content, RNA editing, and gene migration to the nucleus. *Annual Review of Plant Physiology and Plant Molecular Biology*, 45, 61-78.
- [23] Palmer, J. D. (1992). Mitochondrial DNA in plant systematics: applications and limitations. In: Soltis PS, Soltis DE, Doyle JJ. (ed.) *Molecular systematics of plants*, New York, Chapman and Hall, 26-49.
- [24] Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. USA. *Proceedings of the National Academy of Sciences*, 84, 9054-9058.
- [25] Larsson, N. G., & Clayton, D. A. (1995). Molecular genetic aspects of human mitochondrial disease. *Annual Review of Genetics*, 29, 151-178.
- [26] Binder, S., Marchfelder, A., & Brennicke, A. (1996). Regulation of gene expression in plant mitochondria. *Plant Molecular Biology*, 32, 303-314.
- [27] Unseld, M., Marienfeld, J., Brandt, P., & Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics*, 15, 57-61.

- [28] Oda, K., Yamato, K., Ohata, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanrgae, T., Ogura, Y., Kohchi, T., & Ohyama, K. (1992). Gene organization from the complete sequence of liverwort, *Marchantia polymorpha*, mitochondrial DNA: a primitive form of plant mitochondrial genome. *Journal of Molecular Biology*, 223, 1-7.
- [29] Backert, S., Nielsen, B. L., & Borner, T. (1997). The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends in Plant Science*, 2, 477-483.
- [30] Andre, C. P., & Walbot, V. (1995). Pulsed-field gel mapping of maize mitochondrial chromosomes. *Molecular and General Genetics*, 247, 255-263.
- [31] Bendich, A. J. (1993). Reaching for the ring: the study of mitochondrial genome structure. *Current Genetetics*, 24, 279-290.
- [32] Fauron, C., Casper, M., Gao, Y., & Moore, B. (1995). The maize mitochondrial genome: dynamic, yet functional. *Trends in Genetics*, 11, 228-235.
- [33] Fauron, C. M. R., Casper, M., Gesteland, R., & Albertsen, M. (1992). A multi-recombination model for the mtDNA rearrangements seen in maize cmsT regenerated plants. *The Plant Journal*, 2, 949-958.
- [34] Palmer, J. D., & Herbon, L. A. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, 28, 87-97.
- [35] Palmer, J. D., & Shields, C. R. (1984). Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature*, 307, 437-440.
- [36] Palmer, J. D., Makaroff, C. A., Apel, I. J., & Shirzadegan, M. (1990). Fluid structure of plant mitochondrial genomes: evolutionary and functional implications. In: *Clegg MT, O'Brien SJ (ed.) Molecular Evolution*, New York, Alan R. Liss, 85-96.
- [37] Mosig, G. (1987). The essential role of recombination in phage T4 growth. *Annual Review of Genetics*, 21, 347-371.
- [38] Mosig, G. (1998). Recombination and recombinationdependent DNA replication in bacteriophage T4. *Annual Review of Genetics*, 32, 379-413.
- [39] Oldenburg, D. J., & Bendich, A. J. (2001). Mitochondrial DNA from the liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 50 protein. *Journal of Molecular Biology*, 310, 549-562.
- [40] Backert, S., Dorfel, P., & Borner, T. (1995). Investigation of plant organellar DNA by pulsed-field gel electrophoresis. *Current Genetics*, 28, 390-399.
- [41] Backert, S., Lurz, R., Oyarzabal, O. A., & Borner, T. (1997). High content, size and distribution of singlestranded DNA in the mitochondria of *Chenopodium album* (L.). *Plant Molecular Biology*, 33, 1037-1050.

- [42] Bendich, A. J., & Smith, S. B. (1990). Moving pictures and pulsed-field gel electrophoresis show linear DNA molecules from chloroplasts and mitochondria. *Current Genetics*, 17, 421-425.
- [43] Bendich, A.J. (1996). Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. *Journal of Molecular Biology*, 225, 564-588.
- [44] Oldenburg, D. J., & Bendich, A. J. (1996). Size and structure of replicating mitochondrial DNA in cultured tobacco cells. *Plant Cell*, 8, 447-461.
- [45] Sena, E. P., Revet, B., & Moustacchi, E. (1986). In vivo homologous recombination intermediates of yeast mitochondrial DNA analyzed by electron microscopy. *Molecular and General Genetics*, 202, 421-428.
- [46] Palmer, J. D., & Herbon, L. A. (1987). Unicircular structure of the Brassica hirta mitochondrial genome. *Current Genetics*, 11, 565-570.
- [47] Stern, D. B., & Palmer, J. D. (1984). Recombination sequences in plant mitochondrial genomes: Diversity and homologies to known mitochondrial genes. *Nucleic Acid Research*, 12, 6141-6157.
- [48] Chanut, F. A., Grabau, E. A., & Gesteland, R. F. (1993). Complex organization of the soybean mitochondrial genome: recombination repeats and multiple transcripts at the atpA loci. *Current Genetics*, 23, 234-247.
- [49] Dewey, R. E., Levings, C. S., III., & Timothy, D. H. (1986). Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male-sterile cytoplasm. *Cell*, 44, 439.
- [50] Laver, H. K., Reynolds, S. J., Moneger, F., & Leaver, C. J. (1991). Mitochondrial genome organization and expression associated with cytoplasmic male sterility in sunflower (*Helianthus annuus*). *Plant Journal*, 1, 185-193.
- [51] Johns, C., Lu, M., Lyznik, A., & Mackenzie, S. (1994). A mitochondrial DNA sequence is associated with abnormal pollen development in cytoplasmic male sterile bean plants. *Plant Cell*, 4, 435-449.
- [52] Martinez-Zapater, J. M., Gil, P., Capel, J., & Somerville, C. R. (1992). Mutations at the Arabidopsis CHM locus promote rearrangements of the mitochondrial genome. *Plant Cell*, 4, 889-899.
- [53] Newton, K., Knudsen, C., Gabay-Laughnan, S., & Laughnan, J. (1990). An abnormal growth mutant in maize has a defective mitochondrial cytochrome oxidase gene. *Plant Cell*, 2, 107-113.
- [54] Levings, C. S., III., & Pring, D. R. (1979). Mitochondrial DNA of higher plants and genetic engineering. In: Setlow JK, Hollaender A (ed.). *Genetic engineering*, 1, New York, Plenum Press, 205-222.

- [55] Bailey-Serres, J., Leroy, P., Jones, S. S., Wahleithner, J. A., & Wolstenholme, D. R. (1987). Size distribution of circular molecules in plant mitochondrial DNAs. *Current Genetics*, 12, 49-53.
- [56] Grabau, E., Davis, W. H., & Gengenbach, B. G. (1989). Restriction fragment length polymorphism in a subclass of the mandarin soybean. *Crop Science*, 29, 1554-1559.
- [57] Synenki, R. M., Levings, C. S., III., & Shah, D. M. (1978). Physicochemical characterization of mitochondrial DNA from soybean. *Plant Physiology*, 61, 460-464.
- [58] Kanazawa, A., Tozuka, A., Kato, S., Mikami, T., Abe, J., & Shimamoto, Y. (1998). Small interspersed sequences that serve as recombination sites at the *cox2* and *atp6* loci in the mitochondrial genome of soybean are widely distributed in higher plants. *Current Genetics*, 33, 188-198.
- [59] Kato, S., Kanazawa, A., Mikami, T., & Shimamoto, Y. (1998). Evolutionary changes in the structures of the *cox2* and *atp6* loci in the mitochondrial genome of soybean involving recombination across small interspersed sequences. *Current Genetics*, 34, 303-312.
- [60] Grabau, E. A., Davis, W. H., Phelps, N. D., & Gengenbach, B. G. (1992). Classification of soybean cultivars based on mitochondrial DNA restriction fragment length polymorphisms. *Crop Science*, 32, 271-274.
- [61] Moeykens, C. A., Mackenzie, S. A., & Shoemaker, R. C. (1995). Mitochondrial genome diversity in soybean: repeats and rearrangements. *Plant Molecular Biology*, 29, 245-254.
- [62] Hanlon, R. W., & Grabau, E. A. (1997). Comparison of mitochondrial organization of four soybean cytoplasmic types by restriction mapping. *Soybean Genetics Newsletter*, 24, 208-210.
- [63] Khazi, F. R., Edmondson, A. C., & Nielsen, B. L. (2003). An Arabidopsis homologue of bacterial RecA that complements an *E. coli* *recA* deletion is targeted to plant mitochondria. *Molecular and General Genetics*, 269, 454-463.
- [64] Manchekar, M., Scissum-Gunn, K., Song, D., Khazi, F., Mc Lean, S. L., & Nielsen, B. L. (2006). DNA Recombination Activity in Soybean Mitochondria. *Journal of Molecular Biology*, 356, 288-299.
- [65] Grabau, E. A., Havlik, M., & Gesteland, R. F. (1988). Chimeric Organization of Two Genes for the Soybean Mitochondrial Atpase Subunit 6. *Current Genetics*, 13, 83-89.
- [66] Wissinger, B., Brennicke, A., & Schuster, W. (1992). Regenerating good sense: RNA editing and trans-splicing in plant mitochondria. *Trends in Genetics*, 8, 322-328.
- [67] Hanlon, R., & Grabau, E. A. (1995). Cytoplasmic diversity in old domestic varieties of soybean using two mitochondrial markers. *Crop Science*, 35, 1148-1151.
- [68] Grabau, E. A., & Davis, W. H. (1992). Cytoplasmic diversity in *Glycine soja*. *Soybean Genetics Newsletter*, 19, 140-144.

- [69] Tozuka, A., Fukushi, H., Hirata, T., Ohara, M., Kanazawa, A., Mikami, T., Abe, J., & Shimamoto, Y. (1998). Composite and clinal distribution of Glycine soja in Japan revealed by RFLP analysis of mitochondrial. *DNA Theoretical and Applied Genetics*, 96, 170-176.
- [70] Schuster, W., & Brennicke, A. (1987). Nucleotide sequence of the *Oenothera* ATPase subunit 6 gene. *Nucleic Acids Research*, 15, 9092.
- [71] Braun, C. J., & Levings, C. S., III. (1985). Nucleotide Sequence of the F(1)-ATPase alpha Subunit Gene from Maize Mitochondria. *Plant Physiology*, 79(2), 571-577.
- [72] Isaac, G. I., Brennicke, A., Dunbar, M., & Leaver, C. J. (1985). The mitochondrial genome of fertile maize (*Zea mays* L.) contains two copies of the gene encoding the alpha-subunit of the F1-ATPase. *Current Genetics*, 10(4), 321-328.
- [73] Schuster, W., & Brennicke, A. (1986). Pseudocopies of the ATPase a-subunit gene in *Oenothera* mitochondria are present on different circular molecules. *Molecular and General Genetics*, 204, 29-35.
- [74] Morikami, A., & Nakamura, K. (1987). Structure and expression of pea mitochondrial F1ATPase alpha-subunit gene and its pseudogene involved in homologous recombination, *J Biochem*, 101, 967-976.
- [75] Chaumont, F., Boutry, M., Briquet, M., & Vassarotti, A. (1988). Sequence of the gene encoding the mitochondrial F1ATPase alpha subunit from *Nicotiana plumbaginifolia*. *Nucleic Acids Research*, 16(13), 6247.
- [76] Schulte, E., Staubach, S., Laser, B., & Kück, U. (1989). Wheat mitochondrial DNA: organization and sequences of the *atpA* and *atp9* genes. *Nucleic Acids Research*, 17(18), 7531.
- [77] Kadowaki, K., Kazama, S., & Suzuki, S. (1990). Nucleotide sequence of the F₁-ATPase a subunit genes from rice mitochondria. *Nucleic Acids Research*, 18, 1302.
- [78] Köhler, R. H., Lössl, A., & Zetsche, K. (1990). Nucleotide sequence of the F1ATPase alpha subunit gene of sunflower mitochondria. *Nucleic Acids Research*, 18, 4588.
- [79] Makaroff, C. A., Apel, I. J., & Palmer, J. D. (1990). Characterization of radish mitochondrial *atpA*: influence of nuclear background on transcription of *atpA*-associated sequences and relationship with male sterility. *Plant Molecular Biology*, 15(5), 735-746.
- [80] Siculella, L., D'Ambrosio, L., De Tuglie, A. D., & gauerani, R. (1990). Minor differences in the primary structures of *atpA* genes coded on the mt DNA of fertile and male sterile sunflower lines. *Nucleic Acids Research*, 18(15), 4599.
- [81] Handa, H., & Nakajima, K. (1991). Nucleotide sequence and transcription analyses of the rapeseed (*Brassica napus* L.) mitochondrial F1ATPase alpha-subunit gene. *Plant Molecular Biology*, 16, 361-364.

- [82] Chase, C. D., & Ortega, V. M. (1992). Organization of ATPA coding and 3' flanking sequences associated with cytoplasmic male sterility in *Phaseolus vulgaris*. *L. Current Genetics*, 22, 147-153.
- [83] Zurawski, G., Bottomley, W., & Whitfield, P. R. (1982). Structures of the genes for the beta and c subunits of spinach chloroplast ATPase indicates a dicistronic mRNA and an overlapping translation stop/start signal. USA. *Proceedings of the National Academy of Sciences*, 79, 6260-6264.
- [84] Krebbers, E. T., Larrinua, I. M., Mc Intosh, L., & Bogorad, L. (1982). The maize chloroplast genes for the beta and epsilon subunits of the photosynthetic coupling factor CF1 are fused. *Nucleic Acids Research*, 10(16), 4985-5002.
- [85] Walker, J. E., Fearnley, I. M., Gay, N. J., Gibson, B. W., Northrop, F. D., Powell, S. J., Runswick, M. J., Saraste, M., & Tybulewicz, V. L. J. (1985). Primary structure and subunit stoichiometry of F1-ATPase from bovine mitochondria. *Journal of Molecular Biology*, 184(4), 677-701.
- [86] Shimamoto, Y., Fukushi, H., Abe, J., Kanazawa, A., Gai, J., Gao, Z., & Xu, D. (1998). RFLPs of chloroplast and mitochondrial DNA in wild soybean, *Glycine soja*, growing in China. *Genetic Resources and Crop Evolution*, 45, 433-439.
- [87] Abe, J. (1999, 13th-15th October). The genetic structure of natural populations of wild soybeans revealed by isozymes and RFLP's of mitochondrial DNS's: possible influence of seed dispersal, cross pollination and demography. Japan. In: *The Seventh Ministry of Agriculture, Forestry and Fisheries (MAFF), Japan International Workshop on Genetic Resources Part I Wild Legumes*, National Institute of Agrobiological Resources Tsukuba, Ibaraki, 143-159.
- [88] Shimamoto, Y. (2001). Polymorphism and phylogeny of soybean based on chloroplast and mitochondrial DNA analysis. *Japan Agricultural Research Quarterly (JARQ)*, 35(2), 79-84.
- [89] Cleeg, M. T., & Zurawski, G. (1992). Chloroplast DNA and the study of plant phylogeny. New York, In: *Soltis PS, Soltis DE, Doyle JJ (eds.) Present status and future prospects in Molecular Systematics of Plants*, Chapman & Hall, 1-13.
- [90] Olmstead, R. G., & Palmer, J. D. (1994). Chloroplast DNA systematics: a review of method and data analysis. *American Journal of Botany*, 81, 1205-1224.
- [91] Chase, M., & 41others, . (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcl*. *Annals Missouri Botany Gardens*, 80, 528-580.
- [92] Parkinson, C. L., Adams, L., & Palmer, D. (1999). Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology*, 9, 1485-1488.
- [93] Qui, Y. L. (1999). The earliest angiosperms: Evidence from mitochondrial, plastid, and nuclear genomes. *Nature*, 402, 404-407.

- [94] Soltis, D. E., Soltis, P. S., & Doyle, J. J. (1998). *Molecular systematic of Plants II. DNA sequencing*, Boston MA, Kluwer.
- [95] Soltis, P. S., Soltis, D. E., & Chase, M. W. (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature*, 402, 402-404.
- [96] Kato, T., Kaneko, T., Sato, S., Nakamura, Y., & Tabata, S. (2000). Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Research*, 7, 323-330.
- [97] Li, W. H. (1997). *Molecular Evolution*, Sunderland, Sinauer Associates, 487.
- [98] Sasaki, C.h., Lee, S. B., Daniell, H., Wood, T. C., Tomkins, J., Kim, H. G., & Jansen, R. K. (2005). Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Molecular Biology*, 59, 309-322.
- [99] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., & Sugiura, M. (1986). The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. *EMBO Journal*, 5, 2043-2049.
- [100] Sugiura, M., & Shimada, H. (1991). Fine- structural features of the chloroplast genome- comparison of the sequenced chloroplast genomes. *Nucleic Acids Research*, 19, 983-995.
- [101] Sugiura, M., & Sugita, M. (1992). The chloroplast genome. *Plant Molecular Biology*, 19, 149-168.
- [102] Woźny, A. (1990). *Wykłady i ćwiczenia z biologii komórki*, Warszawa, Wydawnictwo Naukowe PWN.
- [103] Palmer, J. D. (1985). Evolution of chloroplast and mitochondrial DNA in plants and algae. In: *MacIntyre R.J. (ed.) Monographs in Evolutionary Biology*, New York., Molecular Evolutionary Genetics Plenum Press, 131-240.
- [104] Wolfe, K.H. (1988). The site of deletion of the inverted repeat in pea chloroplast DNA contains duplicated gene fragments. *Current Genetic*, 13, 97-99.
- [105] Palmer, J. D., Osorio, B., Aldrich, J., & Thompson, W. F. (1987). Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. *Current Genetic*, 11, 275-286.
- [106] Lavin, M., Doyle, J. J., & Palmer, J. D. (1990). Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution*, 44, 390-402.
- [107] Herdenberger, F., Pillay, D. T. N., & Steinmetz, A. (1990). Sequence of the trnH gene and the inverted repeat structure deletion of the broad bean chloroplast genome. *Nucleic Acids Research*, 18, 1297.
- [108] Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E. E., & Small, R. L. (2005). The tortoise and the hare II: rela-

- tive utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, 92, 142-166.
- [109] Ravi, V., Khurana, J. P., Tyagi, A. K., & Khurana, P. (2008). An update on chloroplast genomes. *Plant Systematics and Evolution*, 271101-122.
- [110] Kelchner, S. A. (2000). The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden*, 87, 482-498.
- [111] Perry, A. S., Brennan, S., Murphy, D. J., & Wolfe, K. H. (2006). Evolutionary re-organisation of a large operon in Adzuki bean chloroplast DNA caused by inverted repeat movement. *DNA Research*, 9, 157-162.
- [112] Close, P. S., Shoemaker, R. C., & Keim, P. (1989). Distribution of restriction site polymorphism within the chloroplast genome of the genus *Glycine*, subgenus *Soja*. *Theoretical and Applied Genetics*, 77, 768-776.
- [113] Xu, D. H., Abe, J., Kanazawa, A., Sakai, M., & Shimamoto, Y. (2000). Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. *TAG Theoretical and Applied Genetics*, 101, 5-6, 724-732.
- [114] Kajita, T., Ohashi, H., Tateishi, Y., Bailey, C. D., & Doyle, J. J. (2001). *rbcL* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. *Systematic Botany*, 26, 515-536.
- [115] Nickret, D. L., & Patrick, J. A. (1998). The nuclear ribosomal DNA intergenic spacer of wild and cultivated soybean have low variation and cryptic subrepeats. *NRC Canada. Genome*, 41183-192.
- [116] 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. U.S.A., *Proceedings of the National Academy of Sciences*, 97, 6960-6966.
- [117] Pennington, R. T., Klitgaard, B. B., Ireland, H., & Lavin, M. (2000). New insights into floral evolution of basal Papilionoideae from molecular phylogenies. In: *PS Herendeen A Bruneau (ed.) Advances in Legume Systematics Kew, UK*, 9, 233-248.
- [118] Wojciechowski, M. F., Lavin, M., & Sanderson, M. J. (2004). A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany*, 91, 1846-1862.
- [119] Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, 17, 1105-1109.
- [120] Wolfe, A. D., Elisens, W. J., Watson, L. E., & Depamphilis, C. W. (1997). Using restriction-site variation of PCR-amplified cpDNA genes for phylogenetic analysis of Tribe Cheloneae (Scrophulariaceae). *American Journal of Botany*, 84, 555-564.

- [121] Sehgal, D., & Raina, S. N. (2008). DNA markers and germplasm resource diagnostics: new perspectives in crop improvement and conservation strategies. *In: Arya ID, Arya S. (ed.) Utilization of biotechnology in plant sciences*, Microsoft Printech (I) Pvt. Ltd., Dehradun, 39-54.
- [122] Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnology*, 23, 48-55.
- [123] Varshney, R. K., Thudi, M., Aggarwal, R., & Börner, A. (2007). Genic molecular markers in plants: development and applications. *In: Varshney RK, Tuberosa R (ed.) Genomics assisted crop improvement: genomics approaches and platforms*, Springer, Dordrecht, 1, 13-29.
- [124] Powell, W., Machray, G. C., & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Sciences*, 1, 215-222.
- [125] Russell, J. R., Fuller, J. D., Macaulay, M., Hatz, B. G., Jahoor, A., Powell, W., & Waugh, R. (1997). Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theoretical and Applied Genetics*, 95, 714-722.
- [126] Bohn, M., Utz, H. F., & Melchinger, A. E. (1999). Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Sciences*, 39, 228-237.
- [127] Parida, S. K., Kalia, S. K., Sunita, K., Dalal, V., Hemaprabha, G., Selvi, A., Pandit, A., Singh, A., Gaikwad, K., Sharma, T. R., Srivastava, P. S., Singh, N. K., & Mohapatra, T. (2009). Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. *Theoretical and Applied Genetics*, 118, 327-338.
- [128] Shimamoto, Y., Hasegawa, A., Abe, J., Ohara, M., & Mikami, T. (1992). Glycine soja germplasm in Japan: isozymes and chloroplast DNA variation. *Soybean Genet News-letter*, 19, 73-77.
- [129] Xu, D. H., Abe, J., Gai, J. Y., & Shimamoto, Y. (2002). Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theoretical and Applied Genetics*, 105, 645-653.
- [130] Xu, D. H., Abe, J., Kanazawa, A., Gai, Y., & Shimamoto, Y. (2001). Identification of sequence variations by PCR-RFLP and its application to the evaluation of cpDNA diversity in wild and cultivated soybeans. *Theoretical and Applied Genetics*, 102, 683-688.
- [131] Hirata, T., Abe, J., & Shimamoto, Y. (1996). RFLPs of chloroplast and mitochondrial genomes in summer and autumn maturing cultivar groups of soybean in Kyushu district of Japan. *Soybean Genetic Newsletter*, 23, 107-111.
- [132] Shimamoto, A., Fukushi, H., Abe, J., Kanazawa, A., Gai, J., Gao, Z., & Xu, D. (1998). RFLPs of chloroplast and mitochondrial DNA in wild soybean, *Glycine soja*, growing in China. *Genetic Resources and Crop Evolution*, 45, 433-439.

- [133] Chen, J., & Hebert, P. D. N. (1999). Directed termination of the polymerase chain reaction: kinetics and application in mutation detection. *Genome*, 42, 72-79.
- [134] Orita, M., Suzuki, T. S., & Hayashi, K. (1989). Rapid and sensitive detection of point mutation and DNA polymorphisms using the polymerase chain reaction. *Genomics*, 5, 874-879.
- [135] Sarkar, G., Yoon, H., & Sommer, S. S. (1995). Dideoxy fingerprinting (ddF): a rapid and efficient screen for the presence of mutations. *Genomics*, 13, 441-443.
- [136] Manen, J. F., & Natall, A. (1995). Comparison of the evolution of ribulose- 1, 5-bisphosphate carboxylase (rbcL) and atpB-rbcL non-coding spacer sequences in a recent plant group, the tribe Rubieae (Rubiaceae). *Journal of Molecular Evolution*, 41, 920-927.
- [137] 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). *American Journal of Botany*, 83, 430-439.
- [138] Sang, T., Crawford, D. J., & Stuessy, T. F. (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). *American Journal of Botany*, 84, 1120-1136.
- [139] Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., & Wendel, J. F. (1998). The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany*, 85, 1301-1315.
- [140] Mc Dade, L. A., & Moody, M. L. (1999). Phylogenetic relationship among Acanthaceae: evidence from noncoding trnL-trnFF chloroplast DNA sequences. *American Journal of Botany*, 86, 70-80.
- [141] Molvray, M., Kores, P. J., & Chase, M. W. (1999). Phylogenetic relationships within Korthalsella (Viscaceae) based on nuclear ITS and plastid trnL-F sequence data. *American Journal of Botany*, 86, 249-260.
- [142] Demesure, B., Comps, B., & Petit, R. J. (1996). Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, 50, 2515-2520.
- [143] Wolfe, A. D., Elisens, W. J., Watson, L. E., & Depamphilis, C. W. (1997). Using restriction-site variation of PCR-amplified cpDNA genes for phylogenetic analysis of Tribe Cheloneae (Scrophulariaceae). *American Journal of Botany*, 84, 555-564.
- [144] Asmussen, C. B., & Liston, A. (1998). Chloroplast DNA characters, phylogeny, and classification of Lathyrus (Fabaceae). *American Journal of Botany*, 85, 387-40.
- [145] Cipriani, G., Testolin, R., & Gardner, R. (1998). Restriction-site variation of PCR-amplified chloroplast DNA regions and its implication for the evolution of Actinidia. *Theoretical and Applied Genetics*, 96, 389-396.

- [146] Friesen, N., Pollner, S., Bachmann, K., & Blattner, F. R. (1999). RAPDs and noncoding chloroplast DNA reveal a single origin of the cultivated *Allium fistulosum* from *A. altaicum* (Alliaceae). *American Journal of Botany*, 86, 554-562.
- [147] Close, P. S., Shoemaker, R. C., & Keim, P. (1989). Distribution of restriction site polymorphism within the chloroplast genome of the genus *Glycine*, subgenus *Soja*. *Theoretical and Applied Genetics*, 77, 768-776.
- [148] Kanazawa, A., Tozuka, A., & Shimamoto, Y. (1998). Sequence variation of chloroplast DNA that involves *EcoRI* and *ClaI* restriction site polymorphisms in soybean. *Genes Genetic Systems*, 73, 111-119.
- [149] Downie, S. R., Olmstead, R. G., Zurawski, G., Soltis, D. E., Soltis, P. S., Watson, C., & Palmer, J. D. (1991). Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: Molecular and phylogenetic implications. *Evolutions*, 45, 1245-1259.
- [150] Spielmann, A., & Stutz, E. (1983). Nucleotide sequence of soybean chloroplast DNA regions which contain the *psbA* and *trnH* genes and cover the ends of the large single-copy region and one end of the inverted repeats. *Nucleic Acids Research*, 11, 7157-7167.
- [151] Sugiura, M., & Shimada, H. (1991). Fine- structural features of the chloroplast genome- comparison of the sequenced chloroplast genomes. *Nucleic Acids Research*, 19, 983-995.
- [152] Taberlet, P., Coissac, E., Popanon, F., Gielly, L., Miquel, C., Valentini, A., Vermet, T., Corthier, G., Brochmann, C., & Willerslev, E. (2006). Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, 35(3).
- [153] Palmer, J. D. (1991). Plastid chromosomes: Structure and Evolution. In: *Bogorag L., Vasil IK (ed.) Cell culture and somatic cell genetic of plants*, New York, Academic Press, 7, 5-53.

