DNA Research **17**, 11–22, (2010)
Advance Access Publication: 10 December 2009
doi:10.1093/dnares/dsp025

The Chloroplast Genome Sequence of Mungbean (*Vigna radiata*) Determined by High-throughput Pyrosequencing: Structural Organization and Phylogenetic Relationships

S. Tangphatsornruang*, D. Sangsrakru, J. Chanprasert, P. Uthaipaisanwong, T. Yoocha, N. Jomchai, and S. Tragoonrung

Genome Institute, National Center for Genetic Engineering and Biotechnology, 113 Phaholyothin Rd, Klong 1, Klong Luang, Pathumthani 12120, Thailand

*To whom correspondence should be addressed. Tel. +66 2-564-6700. Fax. +66 2-564-6584.
Email: sithichoke.tan@biotec.or.th

Edited by Katsumi Isono
(Received 11 September 2009; accepted 18 November 2009)

Abstract

Mungbean is an economically important crop which is grown principally for its protein-rich dry seeds. However, genomic research of mungbean has lagged behind other species in the Fabaceae family. Here, we reported the complete chloroplast (cp) genome sequence of mungbean obtained by the 454 pyrosequencing technology. The mungbean cp genome is 151 271 bp in length which includes a pair of inverted repeats (IRs) of 26 474 bp separated by a small single-copy region of 17 427 bp and a large single-copy region of 80 896 bp. The genome contains 108 unique genes and 19 of these genes are duplicated in the IR. Of these, 75 are predicted protein-coding genes, 4 ribosomal RNA genes and 29 tRNA genes. Relative to other plant cp genomes, we observed two distinct rearrangements: a 50-kb inversion between *accD/rps16* and *rbcL/trnK-UUU*, and a 78-kb rearrangement between *trnH/rpl14* and *rps19/rps8*. We detected sequence length polymorphism in the cp homopolymeric regions at the intra- and inter-specific levels in the *Vigna* species. Phylogenetic analysis demonstrated a close relationship between *Vigna* and *Phaseolus* in the phaseolinae subtribe and provided a strong support for a monophyletic group of the eurosid I.

Keywords: mungbean; chloroplast genome; 454 pyrosequencing

1. Introduction

Chloroplasts (cps) are plant organelles that contain biochemical machineries necessary to replicate and transcribe their own genomes. The cp genome of higher plants is a circular molecule of double-stranded DNA and highly conserved in terms of its structure and its gene content with the size ranging from 72 to 217 kb containing ~130 genes, depending on the plant species.[1,2] A pair of large inverted repeats (IRs) that are usually 10–28 kb in length divides the genome into one large single-copy (LSC) region and one small single-copy (SSC) region. Despite high conservation of the cp genome, structural mutations such as gene duplications of tRNA genes,[3] *ycf2*, *rpl19*, *rpl2*, *rpl23*[4] and *psbA*;[5] losses of *ndh* genes,[6] *ycf* genes, *infA* and *accD*;[7–9] as well as rearrangements of cp genomes[10] have been reported in plants and algae. The high level of cp genome rearrangement has been reported in specific lineages including green algae,[11] Campanulaceae,[12] Geraniaceae[13] and Fabaceae.[14]

Mungbean (*Vigna radiata*) is an important economical crop and a member of the Fabaceae family, comprising of around 20 000 species. To date, there are six complete legume cp genomes reported: *Cicer arietinum*[15]

© The Author 2009. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Trifolium subterraneum,[16] *Phaseolus vulgaris*,[17] *Lotus japonicus*,[18] *Glycine max*[19] and *Medicago truncatula* (AC093544, unpublished). Cp genomes of plants in the Fabaceae family are known to have undergone more rearrangements than other angiosperms.[15–20] Palmer and Thompson reported the complete loss of the IR in a group of legume species, such as species from the tribes Carmichaelieae, Cicereae, Hedysareae, Trifolieae, Vicieae and Galegeae.[21–24] Furthermore, a common rearrangement that is shared in most papilionoid legumes is an inversion of a large segment of LSC region relative to all other land plants.[15,20] Within the papilionoid tribe Phaseoleae, another large genome rearrangement in the LSC occurred in subtribe Phaseolinae such as *Vigna* and *Phaseolus*.[14,25] Variations within cp genome sequences are useful for evolutionary studies from population-level processes to more distant phylogenetic relationships.[24] Cp-derived markers, e.g. the *matK* gene and the *trnL-trnF* intergenic spacer, have been used to study the evolutionary relationship between legume plants.[24,26–28] Repetitive sequences within the cp genomes are also potentially useful for ecological and evolutionary studies of plants.[29] A high degree of length polymorphism at cp microsatellite loci has been reported in *Pinus*,[30] *Glycine*,[31] rice[32] and barley.[33] Not only will the information from cp genomes be useful for studies of phylogenetic relationships, but it will also facilitate cp transformation in economically important crops. So far, *G. max* is the only legume in which cp genomes have been successfully transformed to express foreign genes.[34,35]

Since the first report on the complete cp genome of *Marchantia polymorpha*,[36] there are more than 150 complete cp genomes from plants and algae deposited in the GenBank so far. The traditional labour-intensive method for obtaining plastid genome sequence involves isolation of cp DNA followed by random shearing, cloning into BAC or Fosmid vector, and then shotgun sequencing. Recent methods based on long PCR amplification using the conserved cp primers,[37–40] amplification of the entire genome using rolling circle amplification[41,42] and high-throughput sequencing[43–45] have been achieved for fast and cost-effective approaches for cp genome sequencing. In this work, we reported the use of 454 sequencing technology to obtain the genome sequence of mungbean cps for analysis of the structural organization and phylogenetic relationships.

2. Materials and methods

2.1 DNA sequencing, assembly and annotation

The DNA was isolated from 1 g of young leaves of 10-day-old *V. radiata* (L.) Wilczek accession KPS1 using the DNeasy Plant Mini Kit (Qiagen, CA, USA). The DNA (10 μg) was sheared by nebulization, subjected to 454 library preparation and shotgun sequencing using the Genome Sequencer (GS) FLX platform[46] at the in-house facility (National Center for Genetic Engineering and Biotechnology, Thailand). The obtained nucleotide sequence reads were assembled using the Newbler *de novo* sequence assembly software. The cp genome sequence was compared with the reference sequence from the complete cp genome of *P. vulgaris* using the Sequencher software (Gene Codes Corporation, MI, USA). Remaining gaps were closed by PCR and nucleotide sequencing using BigDye Terminator v3.1 Cycle sequencing kit. The primer pairs used for closing the gaps are (i) gap_LSCF: AAT TGG ATA GGA TGG CCT TTG, gap_LSCR: TAG CTC AGT TGG TAG AGC AGA GG; and (ii) gap_IRF: CTG TCC TAG TTG ATC CCG ATT C, gap_IRR: AGA GTG CTT TTT CGA TTC ATC C. For polymorphism test, DNA samples were isolated from young leaves of 10-day-old plants: five samples from *V. radiata* accessions H262 (India), H337 (Afghanistan), H412 (Madagascar), H417 (Nigeria) and KPS1 (Thailand); one sample from *Vigna mungo* accession Subsamotod (Thailand); one sample from *Vigna umbellata* accession JP99485 (Japan); and two samples from *Vigna unguiculata* accessions VU210 (Loas) and TVNU294 (Tanzania), using DNeasy Plant Mini Kit (Qiagen).

2.2 PCR amplification

PCR was carried out in a total volume of 20 μl containing 2 ng of DNA template, 1× buffer, 0.2 mM dNTPs, 1 U Phusion DNA polymerase (Finnzymes, Finland) and 0.5 μM each of forward and reverse primers. The junctions between LSC and IR were confirmed by PCR and nucleotide sequencing using the following primer pairs: (i) JL_F1: GTT TTC AAC AAA ACC CTC TCG T, JL_R1: CCT ACT CTA AAC TTC CGA GGA CA; and (ii) JL_F2: ACT CTA AAC TTC CGA GGA CAT GC, JL_R2: TTT ATC TCT CCA ATT CCC TCG AC. The junctions between SSC and IR were confirmed by PCR and nucleotide sequencing using the following primer pairs: (i) JS_F1: CAG CAA CAA CTG GGT TTA TTA CG, JS_R1: TAC TTT ATT CGT TGG GGC CAT AG; and (ii) JS_F2: CTC TTC CAT CAC CTT GAT ATG TAT G, JS_R2: GGG ACA GCT CAT AAT CTT CAT GT. Amplification was performed in a GeneAmp PCR 9700 System thermocycler (Applied Biosystems, CA, USA) programmed as follows: 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 55°C for 1 min, 72°C for 3 min and a final extension step at 72°C for 10 min. Amplified products were run on either 1% agarose gel or 5% denaturing polyacrylamide gel and visualized by silver staining.

2.3 Genome analysis

The genome was annotated using the program DOGMA (Dual Organellar GenoMe Annotator[47]). The predicted annotations were verified using BLAST similarity search.[48] All genes, rRNAs and tRNAs were identified using the plastid/bacterial genetic code. Comparison of cp genome structures of *Glycine*,[19] *Lotus*,[18] *Medicago* [NC_003119], *Phaseolus*,[17] *Cicer*[15] and *Vigna* (published here) was performed using the Mauve software.[49] REPuter[50] was used to identify and locate direct repeat, IR and reverse complement sequences with $n \geq 30$ bp and a sequence identity $\geq$90%.

2.4 Phylogenetic analysis

A set of 25 protein-coding genes including *matK*, *petA*, *petB*, *petD*, *petG*, *petN*, *psaB*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbN*, *psbT*, *rpoB*, *rpoC1*, *rpoC2*, *rps8*, *rps11*, *rps14* and *ycf3*, from 34 cp genomes representing all lineages of angiosperms, were analysed. These 25 genes are commonly present in all 34 cp genomes and publicly available in the GenBank database. Sequences were aligned using MUSCLE version 3.6,[51] the alignment was edited manually. For maximum likelihood (ML) analysis, RAxML version 7.0[52] was used with the GTR matrix (GTR + Γ model). The local bootstrap probability of each branch was calculated by 100 replications. Phylogenetic analyses using maximum parsimony (MP) method was performed using PAUP version 4.0b10.[53] MP searches included 1000 random addition replicates and a heuristic search using tree bisection and reconnection (TBR) branch swapping with the Multrees option. Bootstrap analysis was performed with 100 replicates with TBR branch swapping. TreeView[54] was used for displaying and printing phylogenetic trees.

3. Results and discussion

3.1 Sequencing results and general features

Sequencing of *V. radiata* genomic DNA was carried out using 454 Life Sciences technology on the GS FLX system. A total of 932 958 quality filtered sequence reads were generated with the average read length of 217 bases covering 202 Mb. Assembly of the nucleotide sequence reads was performed to obtain non-redundant contigs and singlets using the Newbler, a *de novo* sequence assembly software. From the assembly analysis, three contigs were shown to be part of the cp genome by alignment with the *P. vulgaris* cp genome using the Sequencher software. These three contigs were assembled from 48 682 reads (5.22%) and the average genome sequencing depth of each nucleotide on the mungbean cp genome was 61.25×. There were three gaps locating in the non-coding regions in the LSC and IRs with sizes of 28 (at the position 15 597–15 624), 286 (at the position 83 732–84 017) and 286 bp (at the position 148 152–148 437), respectively. The 28-bp gap was a $(TA)_{11}$ repeat between *ndhJ* and *trnF-GAA*. The 286-bp gaps were copies of Ψycf2 present in IRa and IRb. These gaps were probably due to the lack of long sequencing reads that can resolve the locations of four copies of the *ycf2* fragment during sequence assembly. These gaps were closed by PCR and nucleotide sequencing using BigDye Terminator v3.1 Cycle sequencing kit. There were no examples of cp DNA containing non-contiguous sequences that might indicate a nuclear location.

Homopolymer is a stretch of the same nucleotide sequence which has been documented to contribute to technical sequencing errors when using 454 Life Science Technology.[44,55] Throughout the mungbean cp genome, there are 205 homopolymers ($n >$ 7 bp); 73 homopolymers were present in 20 protein-coding genes and 132 homopolymers were present in non-coding regions. Among the protein-coding genes, *ycf1* contains the highest number of homopolymers (22), followed by *rpoC2* (8), *rpoC1* (6), *matK* (5) and *ndhF* (5). The Majority of homopolymers present in the 454 data set was 8 bp long (127), followed by 9 bp long (51). As the length of homopolymer increased, fewer numbers were identified. The longest homopolymer sequences were 13 bp long which only presented twice in the data set. Out of 205 homopolymers, 201 were poly(A/T) and 4 were poly(G/C). To determine the accuracy of these sequences, we performed nucleotide sequencing of all homopolymers that are longer than 7 bp using BigDye Terminator v3.1 Cycle sequencing kit (Supplementary Table S1). Although the homopolymer regions had a deep average sequencing depth of 74.6× and high-quality score of QV 64, we observed that 49 bases in 44 positions out of 1763 homopolymeric bases in 205 positions required correction. Of these, 12 positions were from homopolymers present in coding sequences and the other 32 positions were in non-coding regions. The complete cp genome sequence was reported in the DDBJ/EMBL/GenBank nucleotide sequence database (GQ893027).

The complete cp genome size of mungbean is 151 271 bp including the LSC of 80 896 bp, the SSC of 17 427 bp and a pair of IRs of 26 474 bp each (Fig. 1). The mungbean cp genome size is in range with those from other angiosperms. The IRs span from *rps19* to a portion of *ycf1*. The average AT content of the mungbean cp genome is 64.82%, consistent with the AT content reported for other plant cp

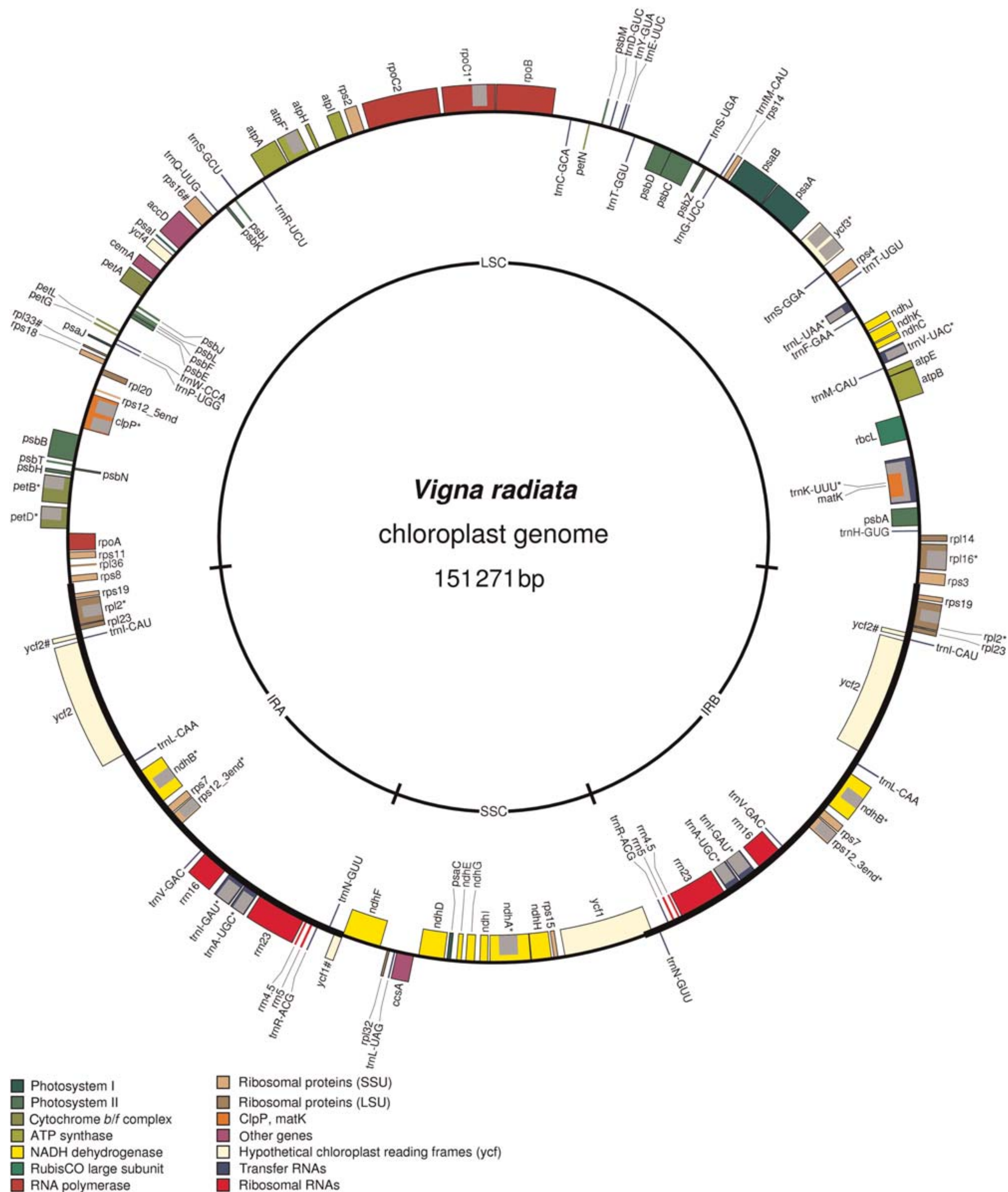

Figure 1. Map of the *V. radiata* cp genome. The thick lines indicate the extent of the IRs (IRa and IRb) which separate the genome into SSC and LSC regions. Genes outside the map are transcribed clockwise and those inside the map are transcribed counter clockwise. Genes containing introns and pseudogenes are marked with * and #, respectively.

genomes. The AT contents of the LSC and SSC regions are 67.4% and 71.35%, respectively, whereas that of the IR regions is 58.26%.

3.2 Genome content and organization

The mungbean cp genome contains 108 unique genes including 29 tRNA genes, 4 rRNA genes and

75 predicted protein-coding genes (Table 1). In addition, there are 19 genes duplicated in the IR, making a total of 127 genes present in the mungbean cp genome. There are 18 genes containing one or more introns, 7 of these are tRNA genes and the other 11 genes are protein-coding genes. The *trnK-UUU* gene has the largest intron (2564 bp) where another gene, the *matK* gene, is present. There are unique 29 tRNA genes (7 tRNA genes duplicated in the IR) representing 20 amino acids identified in the genome (Table 2). On the basis of the sequences of protein-coding genes and tRNA genes within the cp genome, we were able to deduce the frequency of codon usage as summarized in Table 2. We observed that the codon usage was biased towards a high representation of A and T at the third codon position. Non-coding sequences, including intergenic spaces and introns, comprise about 41.45% of the mungbean cp genome, which is close to the proportion of non-coding sequences observed in other cp genomes.[56–60] Two *ycf* genes (*ycf15* and *ycf68*) are probably not functional in the mungbean cp genome due to the presence of premature stop codons. In several cp genomes, *ycf15* and *ycf68* have also been reported as non-functional genes.[17,61,62] A comparison of gene content between the legumes with complete cp genome sequences and the *Arabidopsis* cp genome shows that *rpl22* and *infA* are missing from all legumes.[63,64] Phylogenetic analysis of the nuclear *rpl22* gene suggested the transfer event to the nucleus occurred at an early stage of angiosperm evolution.[64] The loss of *infA* from the cp genome to the nucleus has been reported to occur multiple times during the angiosperm evolution.[9] The *rps16* gene is probably non-functional in *V. radiata* since it contains

Table 1. Genes present in the *V. radiata* cp genome

Gene products
1. Photosystem I: psaA, B, C, I, J, ycf3,[a] ycf4
2. Photosystem II: psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z
3. Cytochrome b6/f: petA, B,[b] D,[b] G, L, N
4. ATP synthase: atpA, B, E, F,[b] H, I
5. Rubisco: rbcL
6. NADH oxidoreductase: ndhA,[b] B,[b,c] C, D, E, F, G, H, I, J, K
7. Large subunit ribosomal proteins: rpl2,[b,c] 14, 16,[b] 20, 23,[c] 32, 36
8. Small subunit ribosomal proteins: rps2, 3, 4, 7,[c] 8, 11, 12,[b–d] 14, 15, 18, 19[c]
9. RNAP: rpoA, rpoB, C1,[b] C2
10. Other proteins: accD, ccsA, cemA, clpP,[a] matK
11. Proteins of unknown function: ycf1, 2[c]
12. Ribosomal RNAs: rrn16,[c] 23,[c] 4.5,[c] 5[c]
13. Transfer RNAs: trnA(UGC),[b,c] C(GCA), D(GUC), E(UUC), F(GAA), G(UCC), H(GUG), I(CAU),[c] I(GAU),[b,c] K(UUU),[b] L(CAA),[c] L(UAA),[b] L(UAG), fM(CAU), M(CAU), N(GUU),[c] P(UGG), Q(UUG), R(ACG),[c] R(UCU), S(GCU), S(GGA), S(UGA), T(GGU), T(UGU), V(GAC),[c] V(UAC),[b] W(CCA), Y(GUA)

[a]Gene containing two introns.
[b]Gene containing a single intron.
[c]Two gene copies in the IRs.
[d]Gene divided into two independent transcription units.

Table 2. The codon–anticodon recognition pattern and codon usage for *L. japonicus* cp genome

Phe	UUU	1125	*trnF-GAA*	Ser	UCU	588	*trnS-GGA*	Tyr	UAU	843	*trnY-GUA*	Cys	UGU	220	*trnC-GCA*
Phe	UUC	517		Ser	UCC	299		Tyr	UAC	158		Cys	UGC	82	
Leu	UUA	929	*trnL-UAA*	Ser	UCA	438	*trnS-UGA*	stop	UAA	53		stop	UGA	15	
Leu	UUG	548	*trnL-CAA*	Ser	UCG	174		stop	UAG	20		Trp	UGG	455	*trnW-CCA*
Leu	CUU	583	*trnL-UAG*	Pro	CCU	417	*trnP-UGG*	His	CAU	496	*trnH-GUG*	Arg	CGU	342	*trnR-ACG*
Leu	CUC	159		Pro	CCC	181		His	CAC	122		Arg	CGC	79	
Leu	CUA	389		Pro	CCA	316		Gln	CAA	756	*trnQ-UUG*	Arg	CGA	351	
Leu	CUG	149		Pro	CCG	128		Gln	CAG	201		Arg	CGG	88	
Ile	AUU	1175	*trnI-GAU*	Thr	ACU	572	*trnT-GGU*	Asn	AAU	1057	*trnN-GUU*	Ser	AGU	401	*trnS-GCU*
Ile	AUC	401		Thr	ACC	199		Asn	AAC	280		Ser	AGC	110	
Ile	AUA	829	*trnI-CAU*	Thr	ACA	433	*trnT-UGU*	Lys	AAA	1243	*trnK-UUU*	Arg	AGA	471	*trnR-UCU*
Met	AUG	585	*trnM-CAU* *trnfM-CAU*	Thr	ACG	126		Lys	AAG	332		Arg	AGG	150	
Val	GUU	528	*trnV-GAC*	Ala	GCU	619	*trnA-UGC*	Asp	GAU	828	*trnD-GUC*	Gly	GGU	594	*trnG-GCC*
Val	GUC	127		Ala	GCC	189		Asp	GAC	196		Gly	GGC	157	
Val	GUA	518	*trnV-UAC*	Ala	GCA	381		Glu	GAA	1017	*trnE-UUC*	Gly	GGA	709	*trnG-UCC*
Val	GUG	168		Ala	GCG	106		Glu	GAG	298		Gly	GGG	254	

Numerals indicate the frequency of usage of each codon in 26 274 codons in 82 potential protein-coding genes.

three internal stop codons and its initial start codon is AGA. The *rps16* gene is lost in other legumes such as *Medicago* and *Cicer*. On the basis of hybridization experiments using *rps16* probe, at least 15 independent losses occurred in the tribe Phaseoleae alone.[65] The mungbean *rpl33* gene contains premature stop codons within the coding region and probably present as a pseudogene as in *Phaseolus*.[17] Instead of a common ATG at the translation initiation site, the mungbean *ndhD* gene has an ACG codon which has also been observed in the cp genomes of *Phaseolus*, *Lotus*, *Pisum*, *Glycine* and *Cicer*. For *ndhD* transcripts, the ACG codon has been shown to convert to the AUG initiation site as reported in pea,[66] leek,[67] tobacco, spinach and snapdragon.[68] It is likely that a similar role of RNA editing in translation process also occurs in mungbean cps.

The cp genome structures of previously sequenced legumes (*Glycine*,[19] *Lotus*,[18] *Medicago* [NC_003119], *Phaseolus*[17] and *Cicer*[15]) and *Vigna* as reported here were compared with the *Arabidopsis* cp genome as the reference sequence using the Mauve software[49] (Fig. 2). Cp genomes of *Medicago* and *Cicer* have lost one copy of the IR and were grouped together in the IR-lacking clade. In mungbean, the border position between the IRa and LSC (JLA) is located in the intergenic region between *rps19* and *rps8*, whereas the junction position between the IRb and LSC (JLB) is located in the intergenic region between *rps19* and *rps3*. The locations of JLA and JLB in mungbean are similar to those in adzuki bean and common bean.[14] In contrast, *Glycine* and *Lotus* cp genomes contain a small fragment of *rps19* in the IR; therefore, *rps19* is only present in the *S10B* Operon (on the JLB side).[18] Shifts in the border positions between IRs and LSC at the JLA and JLB have been reported in several species of angiosperms, demonstrating that the IR/LSC boundaries are dynamic.

The mungbean cp genome, as in common with the *Phaseolus* cp genome, possesses two distinct rearrangements, a 50-kb inversion between *accD*/*rps16* and *rbcL*/*trnK-UUU*, and a 78-kb rearrangement between *trnH*/*rpl14* and *rps19*/*rps8* (Fig. 2). The first inversion is common in papilionoid

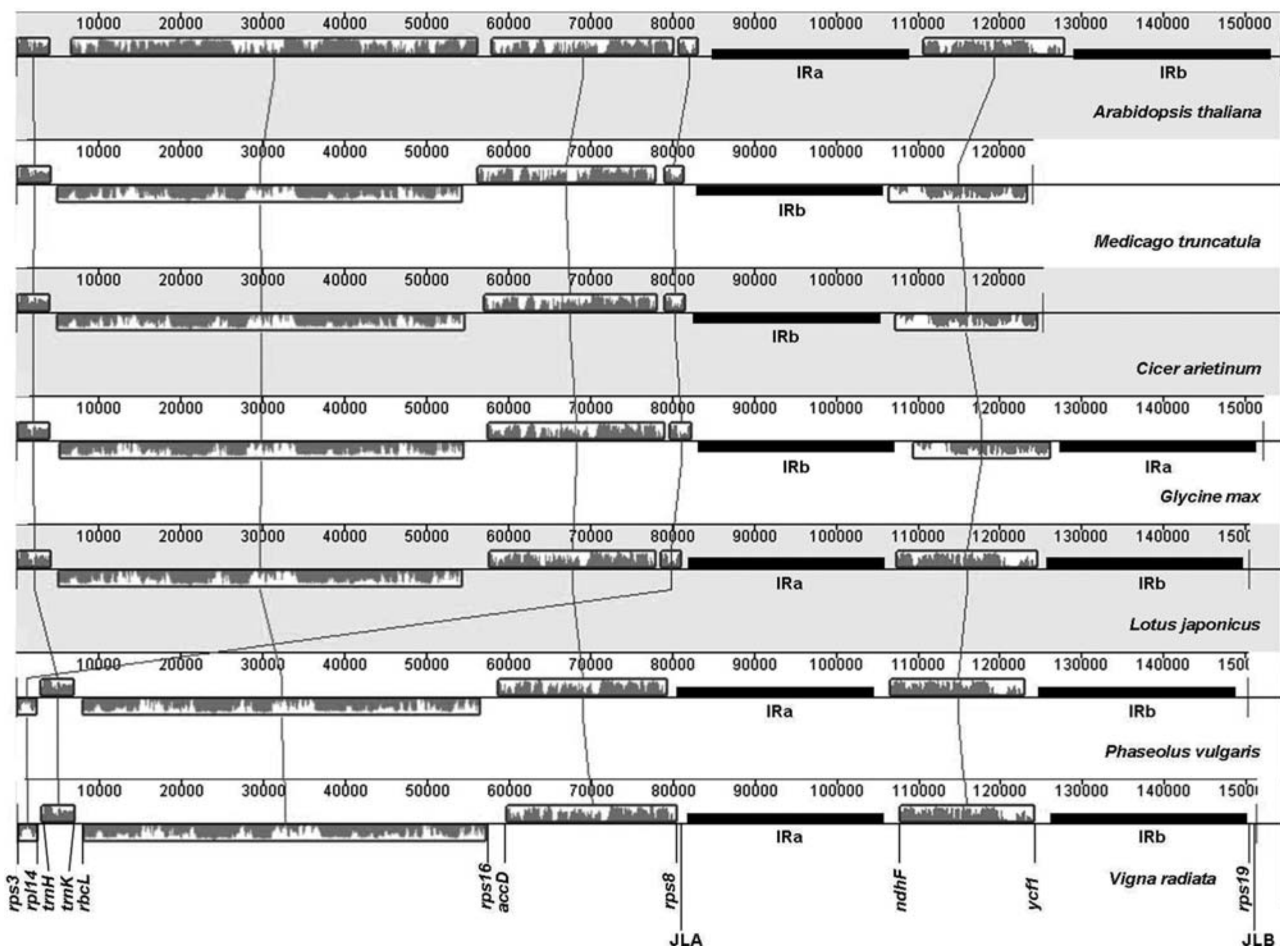

Figure 2. Comparison of legume cp genomes with *Arabidopsis* cp DNA as a reference using MAUVE. The boxes above the line represent DNA sequences in clockwise direction and those below the line represent DNA sequences in the anticlockwise direction. The gene names at the bottom indicate the boundaries of the boxes within the mungbean cp DNA.

legumes indicating an early split in the diversification of papilionoid members.[15,69,70] The second inversion is a distinct rearrangement which is found in subtribe Phaseolinae.[14,20,25] It encompasses nearly the entire fragment of the LSC and disrupts the *S10* operon. Perry *et al.*[14] proposed that this rearrangement occurred after an expansion of the IR such that a copy of the *rpl23-rpl2-rps19-rps3-rpl16-rpl14* genes were introduced in the IRs, followed by the deletion of original genes that became duplicated. Like adzuki bean and *Phaseolus*, the position IR borders of mungbean have expanded to include the entire *rps19* gene when compared with the soybean IR borders. The expansion–contraction model suggested that the contraction did not trim the IR back to its original point as seen in soybean, but instead duplicated the entire *rps19*, leaving two copies in the IR.[14] Therefore, evidence in the mungbean IR border also supports the expansion/contraction mechanism of this 78-kb rearrangement.

3.3 *Polymorphism test of the homopolymeric regions and repeated sequence analysis*

We tested for sequence length polymorphism of 16 homopolymers among five accessions of *V. radiata*, two accessions of *V. unguiculata* (VU210 and TVNU294), one accession of *V. mungo* and one accession of *V. umbellata* (Supplementary Table S2 for details of the primers used, product size and distribution of polymorphism). We observed sequence length polymorphism in 15 loci among *Vigna* species. An example of polymorphic loci is illustrated in Fig. 3. These 15 primer pairs were able to detect a range of 2–4 alleles with a mean of 2.133 alleles per locus. The joint distribution of length variants at these 15 polymorphic loci in the cp DNA revealed five haplotypes among the four *Vigna* species (Supplementary Table S2). Although there was no polymorphism detected between varieties of *V. radiata*, interestingly, we observed polymorphism at the intra-specific level in *V. unguiculata*. This demonstrated that cp microsatellites reported in this study could provide an assay for detecting polymorphism at the population-level and for comparison of more distant phylogenetic relationships at the genus level or above. These cp microsatellites can also be useful in ecological and evolutionary studies because they are non-recombinant, haploid and uniparentally inherited.

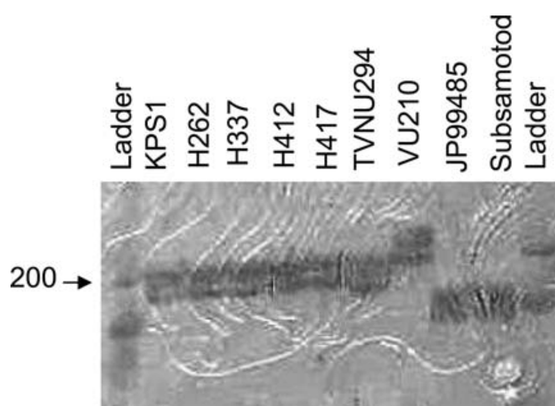

Figure 3. A representative microsatellite locus (Locus 113; forward primer 5′-GAA ACC CTT CCT GAA AAA TCC-3′ and reverse primer 5′-TCT TTG ACG AAT GCA AGT GG-3′) with polymorphism among *Vigna* species: *V. radiata* accessions KPS1, H262, H337, H412, H417, *V. mungo* accession Subsamotod, *V. umbellata* accession JP99485, *V. unguiculata* accessions VU210 and TVNU294. PCR products were separated on 5% polyacrylamide gel electrophoresis and visualized by silver staining.

Analysis of the repeat sequences in the mungbean cp genome identified 22 direct repeats and 28 IRs of 30 bp or longer with a sequence identity of 90% (Table 3). Thirty repeats are 30–40 bp long, 11 repeats are 41–50 bp long, 4 repeats are 51–80 bp long and 5 repeats are longer than 80 bp. The longest direct repeat in mungbean cp DNA is a 287-bp duplication of an internal fragment of *ycf2* ($_\psi$ycf2) in the IRs which shared a very high sequence homology with those of *G. max* and *P. vulgaris*. Most of the direct repeats are distributed within the intergenic spacer regions, the intron sequences, and in the *trnS*, and *ycf2* genes.

3.4 *Phylogenetic analysis*

Our phylogenetic data set included 25 protein-coding genes for 34 plant taxa (Supplementary Table S3), including 32 angiosperms and two outgroup gymnosperms (*Pinus* and *Ginkgo*). These 25 genes are commonly present in all the 34 cp genomes, therefore should reduce missing data from the sequence alignment. The sequence alignment that was used for phylogenetic analyses comprised 20 454 characters. MP analysis resulted in a single resolved tree with a length of 29 081, a consistency index of 0.4939 and a retention index of 0.6399 (Fig. 4). Bootstrap analyses indicated that there were 27 out of 31 nodes with values ≥95%, and 25 of these had a bootstrap value of 100%. ML analysis resulted in a single tree with $-\ln L = -166\,999.997$. ML bootstrap values were also high, with values of ≥95% for 29 of the 32 nodes, and 27 nodes with 100% bootstrap support. Both MP and ML trees had the same topologies which formed two major clades, monocots and eudicots. The trees revealed a monophyly of the monocots and eudicots where Ranunculales was placed as sister to the remaining eudicots. Within the eudicots, there were two major clades: rosids and asterids. Within the

Table 3. A list of repeated sequences and their locations identified using REPuter with $n \geq 30$ bp, and a sequence identity $\geq 90\%$ in the *V. radiata* cp genome

Number	Size (bp)	Repeat	Location
1	42	D	rpl16 (intron):IGS (trnV-GAC and rps12_3end)
2	42	IR	rpl16 (intron): IGS rps12_3end and trnV-GAC
3	50	IR	rpl16 (intron): ndhA (intron)
4	37	D	rpl16 (intron):ycf3 (intron)
5	40	IR	IGS (trnK-UUU and rbcL)
6	81	D	IGS (trnK-UUU and rbcL)
7	40	D	IGS (ndhJ and trnF-GAA)
8	30	D	IGS (ndhJ and trnF-GAA), IGS (psbM and petN)
9	31	IR	trnT-UGU, IGS (psbD and trnT-GGU)
10	31	IR	trnS-GGA, trnS-GCU
11	32	IR	ycf3 (intron)
12	37	IR	ycf3 (intron): ndhA (intron)
13	30	IR	IGS (trnG-UCC and psbZ)
14	31	D	IGS (trnG-UCC and psbZ), IGS (petN and trnC-GCA)
15	31	D	trnS-UGA, trnS-GCU
16	41	IR	IGS (trnG-UCC and psbZ)
17	36	IR	IGS (psbD and trnT-GGU)
18	50	IR	IGS (psbD and trnT-GGU)
19	34	D	IGS (trnC-GCA and rpoB)
20	30	D	IGS (atpI and atpF), IGS (trnL-CAA and ndhB)
21	30	IR	IGS (atpI and atpH), IGS (ndhB and trnL-CAA)
22	30	IR	IGS (trnR-UCU and trnS-GCU)
23	30	IR	IGS (aacD and psaI)
24	34	D	IGS (trnW-CCA and trnP-UGG)
25	41	IR	IGS (psaJ and rpl33_pseudo)
26	50	IR	IGS (psaJ and rpl33_pseudo)
27	80	D	IGS (rps8 and rps19)
28	80	IR	IGS (rps8 and rps19), IGS (rps19 and rps3)
29	287	D	ycf2_pseudo, ycf2
30	287	IR	ycf2_pseudo, ycf2
31	287	IR	ycf2, ycf2_pseudo
32	30	D	ycf2
33	30	IR	ycf2
34	30	IR	ycf2
35	30	D	IGS (trnL-CAA and ndhB)
36	40	D	IGS (rps12_3end and trnV-GAC), ndhA (intron)
37	30	D	IGS (rrn16 and trnI-GAU), IGS (trnI-GAU and rrn16)
38	30	IR	IGS (rrn16 and trnI-GAU)
39	30	D	IGS rrn16 and trnI-GAU: IGS trnI-GAU and rrn16
40	43	D	IGS rrn5 and trnR-ACG: IGS trnR-ACG and rrn5
41	43	IR	IGS (rrn5 and trnR-ACG)
42	31	IR	ndhF
43	40	IR	ndhA (intron): IGS (trnV-GAC and rps12_3end)
44	50	IR	IGS (rps15 and ycf1)
45	43	IR	rrn5
46	30	IR	IGS (trnI-GAU and rrn16)
47	30	D	IGS ndhB and trnL-CAA
48	52	D	ycf2
49	287	D	ycf2:ycf2_pseudo
50	80	D	IGS rps19 and rps3

D, direct repeat; IR, inverted repeat; IGS, intergenic space.

rosid clade, there were two major groups, the eurosids I and eurosids II which were sister to the Myrtales group. The placement of *Cucumis* has been problematic in previous reports.[45,59] In some studies, *Cucumis* was placed with the Myrtales,[45] or in the eurosids I.[58,71,72] In our study, both MP and ML trees provided a strong support for the monophyly of the eurosids I clade because *Cucumis* is sister to the legume taxa. Among seven legumes with complete cp DNA sequences, *Cicer*, *Medicago* and *Trifolium* were grouped together as IR lacking millettioids clade (IRLC) and placed as sister to *Lotus*. IRLC members have been shown to form a monophyletic group supported by phylogenetic trees based on *matK*[24] and nuclear rDNA sequences.[73] *Vigna* was sister to *Phaseolus* in the Phaseolinae subtribe and was sister to *Glycine* in the tribe Phaseoleae. A monophyletic group of the Phaseolinae subtribe was also supported by previously reported trees, based on the *matK* gene[24] and a distinct 78-kb rearrangement of the cp genome.[25]

In conclusion, we performed shotgun genome sequencing of *V. radiata* using the 454 pyrosequencing technology and obtained the complete cp genome sequence. The approach has been demonstrated here as a fast and efficient way for obtaining organellar genomes. Gene content and structural organization of mungbean cp genome is similar to that of *P. vulgaris*, its relative in Phaseoleae. We determined the distribution and the location of repeated sequences in the *V. radiata* cp genome and explored the use of polymorphic microsatellites at the intra-and inter-specific

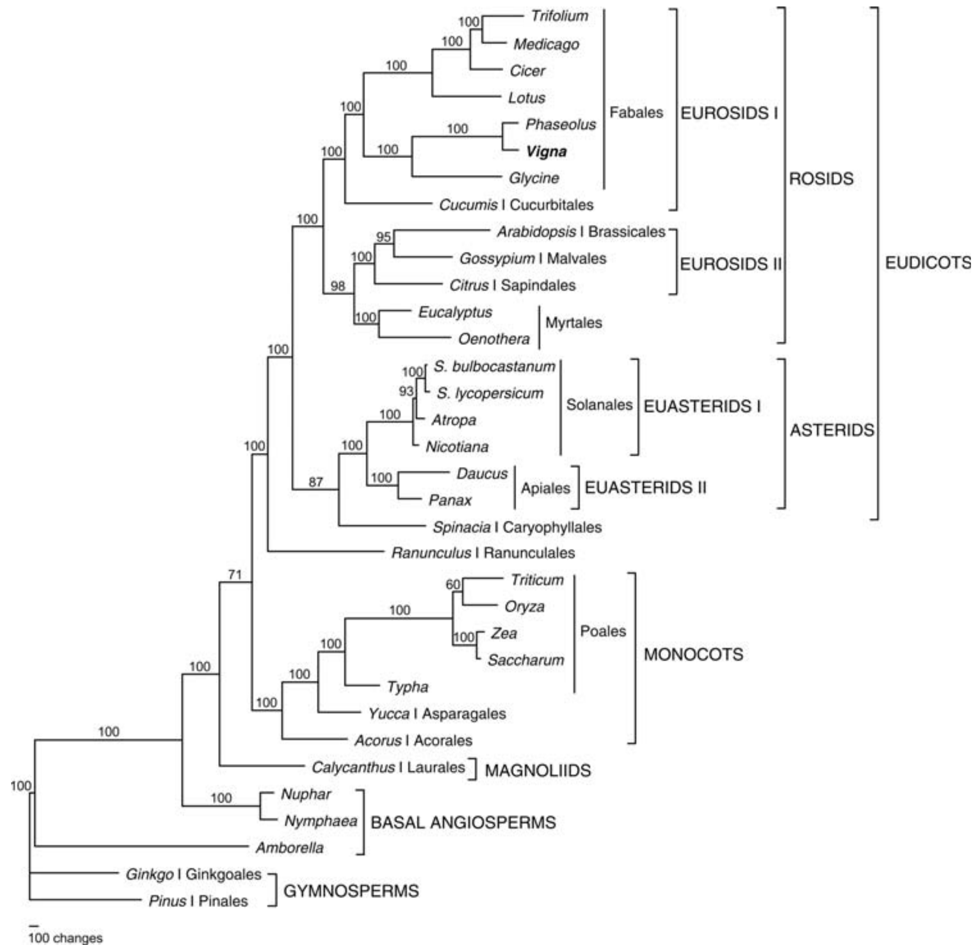

Figure 4. The MP phylogenetic tree is based on 25 protein-coding genes from 34 plant taxa. The MP tree has a length of 29 081 with a consistency index of 0.4939 and a retention index of 0.6399. Numbers above node are bootstrap support values. Ordinal and higher level group names are also indicated. The ML tree has the same topology but is not shown.

levels among *Vigna* species. The proposed phylogenetic relationships among angiosperms, based on cp DNA sequences including those of mungbean cp DNA reported here, provided a strong support for a monophyletic group of the eurosid I and demonstrated a close relationship between *Vigna* and *Phaseolus* in the Phaseolinae subtribe.

Acknowledgements: We are thankful to Prof. Srinives of Kasetsart University, Thailand, for providing samples of *Vigna* species used in this study and Dr Darin Kongkasuriyachai for reviewing the manuscript.

Supplementary Data: Supplementary Data is available at www.dnaresearch.oxfordjournals.org.

Funding

We acknowledge funding support by the Genome Institute, National Center for Genetic Engineering and Biotechnology.

References

1. Sugiura, M. 1995, The chloroplast genome, *Essays Biochem*, **30**, 49–57.
2. Sugiura, M. 1992, The chloroplast genome, *Plant Mol Biol*, **19** (1), 149–68.
3. Hipkins, V.D., Marshall, K.A., Neale, D.B., et al. 1995, A mutation hotspot in the chloroplast genome of a conifer (Douglas-fir: Pseudotsuga) is caused by variability in the number of direct repeats derived from a partially duplicated tRNA gene, *Curr Genet*, **27** (6), 572–9.
4. Bowman, C.M., Barker, R.F. and Dyer, T.A. 1988, In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination, *Curr Genet*, **14** (2), 127–36.
5. Lidholm, J., Szmidt, A. and Gustafsson, P. 1991, Duplication of the psbA gene in the chloroplast genome of two Pinus species, *Mol Gen Genet*, **226** (3), 345–52.
6. Wakasugi, T., Tsudzuki, J., Ito, S., et al. 1994, Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*, *Proc Natl Acad Sci USA*, **91** (21), 9794–8.
7. Maier, R.M., Neckermann, K., Igloi, G.L., et al. 1995, Complete sequence of the maize chloroplast genome:

gene content, hotspots of divergence and fine tuning of genetic information by transcript editing, *J Mol Biol*, **251** (5), 614–28.
8. Hiratsuka, J., Shimada, H., Whittier, R., et al. 1989, The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals, *Mol Gen Genet*, **217** (2–3), 185–94.
9. Millen, R.S., Olmstead, R.G., Adams, K.L., et al. 2001, Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus, *Plant Cell*, **13** (3), 645–58.
10. Wolfe, K.H., Morden, C.W. and Palmer, J.D. 1991, Ins and outs of plastid genome evolution, *Curr Opin Genet Dev*, **1** (4), 523–9.
11. Maul, J.E., Lilly, J.W., Cui, L., et al. 2002, The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats, *Plant Cell*, **14** (11), 2659–79.
12. Cosner, M.E., Raubeson, L.A. and Jansen, R.K. 2004, Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes, *BMC Evol Biol*, **4**, 27.
13. Guisinger, M.M., Kuehl, J.V., Boore, J.L., et al. 2008, Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions, *Proc Natl Acad Sci USA*, **105** (47), 18424–9.
14. Perry, A.S., Brennan, S., Murphy, D.J., et al. 2002, Evolutionary re-organisation of a large operon in adzuki bean chloroplast DNA caused by inverted repeat movement, *DNA Res*, **9** (5), 157–62.
15. Jansen, R.K., Wojciechowski, M.F., Sanniyasi, E., et al. 2008, Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae), *Mol Phylogenet Evol*, **48** (3), 1204–17.
16. Cai, Z., Guisinger, M., Kim, H.G., et al. 2008, Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions, *J Mol Evol*, **67** (6), 696–704.
17. Guo, X., Castillo-Ramirez, S., Gonzalez, V., et al. 2007, Rapid evolutionary change of common bean (*Phaseolus vulgaris* L) plastome, and the genomic diversification of legume chloroplasts, *BMC Genomics*, **8**, 228.
18. Kato, T., Kaneko, T., Sato, S., et al. 2000, Complete structure of the chloroplast genome of a legume, *Lotus japonicus*, *DNA Res*, **7** (6), 323–30.
19. Saski, C., Lee, S.B., Daniell, H., et al. 2005, Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes, *Plant Mol Biol*, **59**, 309–22.
20. Palmer, J.D., Osorio, B. and Thompson, W.F. 1988, Evolutionary significance of inversions in legume chloroplast DNAs, *Curr Genet*, **14**, 65–74.
21. Palmer, J.D., Osorio, B., Aldrich, J., et al. 1987, Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements, *Curr Genet*, **11**, 275–86.
22. Lavin, M., Doly, J. and Palmer, J.D. 1990, Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae, *Evolution*, **44**, 390–402.
23. Palmer, J.D. and Thompson, W.F. 1982, Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost, *Cell*, **29** (2), 537–50.
24. Wojciechowski, M.F., Lavin, M. and Sanderson, M.J. 2004, A phylogeny of legume (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family, *Am J Bot*, **91**, 1846–62.
25. Bruneau, A., Doyle, J. and Palmer, J.D. 1990, A chloroplast DNA inversion as a subtribal character in the Phaseoleae (Leguminosae), *Syst Bot*, **15**, 378–86.
26. Pardo, C., Cubas, P. and Tahiri, H. 2004, Molecular phylogeny and systematics of Genista (Leguminosae) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL-trnF intergenic spacer), *Plant Syst Evol*, **244**, 93–119.
27. Lavin, M., Herendeen, P.S. and Wojciechowski, M.F. 2005, Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary, *Syst Biol*, **54**, 575–94.
28. Hilu, K.W., Borsch, T., Muller, K., et al. 2003, Angiosperm phylogeny based on matK sequence information, *Am J Bot*, **90**, 1758–76.
29. Provan, J., Powell, W. and Hollingsworth, P.M. 2001, Chloroplast microsatellites: new tools for studies in plant ecology and evolution, *Trends Ecol Evol*, **16**, 142–7.
30. Powell, W., Morgante, M., McDevitt, R., et al. 1995, Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines, *Proc Natl Acad Sci USA*, **92** (17), 7759–63.
31. Powell, W., Morgante, M., Andre, C., et al. 1995, Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome, *Curr Biol*, **5** (9), 1023–9.
32. Provan, J., Corbett, G., McNicol, J.W., et al. 1997, Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats, *Genome*, **40** (1), 104–10.
33. Provan, J., Russell, J.R., Booth, A., et al. 1999, Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus Hordeum, *Mol Ecol*, **8** (3), 505–11.
34. Dufourmantel, N., Tissot, G., Goutorbe, F., et al. 2005, Generation and analysis of soybean plastid transformants expressing *Bacillus thuringiensis* Cry1Ab protoxin, *Plant Mol Biol*, **58** (5), 659–68.
35. Dufourmantel, N., Dubald, M., Matringe, M., et al. 2007, Generation and characterization of soybean and marker-free tobacco plastid transformants over-expressing a bacterial 4-hydroxyphenylpyruvate dioxygenase which provides strong herbicide tolerance, *Plant Biotechnol J*, **5** (1), 118–33.
36. Ohyama, K., Fukuzawa, H., Kohchi, T., et al. 1986, Chloroplast gene organization deduced from complete

sequence of liverwort *Marchantia polymorpha* chloroplast DNA, *Nature*, **322**, 572–4.
37. Goremykin, V.V., Hirsch-Ernst, K.I., Wolfl, S., et al. 2003, Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that amborella is not a basal angiosperm, *Mol Biol Evol*, **20** (9), 1499–505.
38. Heinze, B. 2007, A database of PCR primers for the chloroplast genomes of higher plants, *Plant Methods*, **3** (4); doi:10.1186/1746-4811-3-4.
39. Dhingra, A. and Folta, K.M. 2005, ASAP: Amplification, sequencing and annotation of plastomes, *BMC Genomics*, **6**, 176.
40. Goremykin, V.V., Hirsch-Ernst, K.I., Wolfl, S., et al. 2004, The chloroplast genome of *Nymphaea alba*: whole-genome analyses and the problem of identifying the most basal angiosperm, *Mol Biol Evol*, **21** (7), 1445–54.
41. Jansen, R.K., Raubeson, L.A., Boore, J.L., et al. 2005, Methods for obtaining and analyzing whole loroplast genome sequences, *Methods Enzymol*, **395**, 348–84.
42. Bausher, M.G., Singh, N.D., Lee, S.B., et al. 2006, The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms, *BMC Plant Biol*, **6**.
43. Cronn, R., Liston, A., Parks, M., et al. 2008, Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology, *Nucleic Acids Res*, **36** (19), e122.
44. Moore, M.J., Dhingra, A., Soltis, P.S., et al. 2006, Rapid and accurate pyrosequencing of angiosperm plastid genomes, *BMC Plant Biol*, **6**, 17.
45. Cai, Z., Penaflor, C., Kuehl, J.V., et al. 2006, Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids, *BMC Evol Biol*, **6**, 77.
46. Margulies, M., Egholm, M., Altman, W.E., et al. 2005, Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, **437** (7057), 376–80.
47. Wyman, S.K., Jansen, R.K. and Boore, J.L. 2004, Automatic annotation of organellar genomes with DOGMA, *Bioinformatics*, **20** (17), 3252–5.
48. Altschul, S.F., Gish, W., Miller, W., et al. 1990, Basic local alignment search tool, *J Mol Biol*, **215** (3), 403–10.
49. Darling, A.C., Mau, B., Blattner, F.R., et al. 2004, Mauve: multiple alignment of conserved genomic sequence with rearrangements, *Genome Res*, **14** (7), 1394–403.
50. Kurtz, S. and Schleiermacher, C. 1999, REPuter: fast computation of maximal repeats in complete genomes, *Bioinformatics*, **15** (5), 426–7.
51. Edgar, R.C. 2004, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res*, **32** (5), 1792–7.
52. Stamatakis, A. 2006, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, *Bioinformatics*, **22** (21), 2688–90.
53. Swofford, D.L. 2003, *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*, Senauer Associates: Sunderland, MA.
54. Page, R.D. 1996, TreeView: an application to display phylogenetic trees on personal computers, *Comput Appl Biosci*, **12** (4), 357–8.
55. Huse, S.M., Huber, J.A., Morrison, H.G., et al. 2007, Accuracy and quality of massively parallel DNA pyrosequencing, *Genome Biol*, **8** (7), R143.
56. Ruhlman, T., Lee, S.B., Jansen, R.K., et al. 2006, Complete plastid genome sequence of Daucus carota: Implications for biotechnology and phylogeny of angiosperms, *BMC Genomics*, **7**.
57. Jansen, R.K., Kaittanis, C., Saski, C., et al. 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.
58. Lee, S.B., Kaittanis, C., Jansen, R.K., et al. 2006, The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms, *BMC Genomics*, **7**, 61.
59. Daniell, H., Lee, S.B., Grevich, J., et al. 2006, Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes, *Theor Appl Genet*, **112** (8), 1503–18.
60. Kim, J.S., Jung, J.D., Lee, J.A., et al. 2006, Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome, *Plant Cell Rep*, **25** (4), 334–40.
61. Steane, D.A. 2005, Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae), *DNA Res*, **12** (3), 215–20.
62. Schmitz-Linneweber, C., Maier, R.M., Alcaraz, J.P., et al. 2001, The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization, *Plant Mol Biol*, **45** (3), 307–15.
63. Spielmann, A., Roux, E., von Allmen, J.M., et al. 1988, The soybean chloroplast genome: complete sequence of the rps19 gene, including flanking parts containing exon 2 of rpl2 (upstream), but rpl22 (downstream), *Nucleic Acids Res*, **16** (3), 1199.
64. Gantt, J.S., Baldauf, S.L., Calie, P.J., et al. 1991, Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron, *EMBO J*, **10** (10), 3073–8.
65. Doyle, J.J., Doyle, J.L. and Palmer, J.D. 1995, Multiple independent losses of two genes and one intron from legume chloroplast genomes, *Syst Bot*, **20**, 272–94.
66. Inada, M., Sasaki, T., Yukawa, M., et al. 2004, A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence, *Plant Cell Physiol*, **45** (11), 1615–22.
67. Del Campo, E.M., Sabater, B. and Martin, M. 2002, Post-transcriptional control of chloroplast gene expression. Accumulation of stable psaC mRNA is due to downstream RNA cleavages in the ndhD gene, *J Biol Chem*, **277** (39), 36457–64.
68. Neckermann, K., Zeltz, P., Igloi, G.L., et al. 1994, The role of RNA editing in conservation of start codons in chloroplast genomes, *Gene*, **146** (2), 177–82.
69. Palmer, J.D. and Thompson, W.F. 1981, Rearrangements in the chloroplast genomes of mung bean and pea, *Proc Natl Acad Sci USA*, **78** (9), 5533–7.

70. Doyle, J.J., Doyle, J.L., Ballenger, J.A., et al. 1996, The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae, *Mol Phylogenet Evol*, **5** (2), 429–38.
71. Logacheva, M.D., Samigullin, T.H., Dhingra, A., et al. 2008, Comparative chloroplast genomics and phylogenetics of *Fagopyrum esculentum* ssp. *ancestrale*—a wild ancestor of cultivated buckwheat, *BMC Plant Biol*, **8**, 59.
72. Daniell, H., Wurdack, K.J., Kanagaraj, A., et al. 2008, The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron, *Theor Appl Genet*, **116** (5), 723–37.
73. Hu, J.-M., Lavin, M., Wojciechowski, M.F., et al. 2002, Phylogenetic analysis of nuclear ribosomal ITS/5.8 S sequences in the tribe Millettieae (Fabaceae): Poecilanthe-Cyclolobium, the core Millettieae, and the Callerya group, *Syst Bot*, **27**, 722–33.