Kasetsart J. (Nat. Sci.) 45 : 461 - 472 (2011)

Sequence Analysis of the *16SrRNA-rps12* Inverted Repeat Region in Chloroplast DNA of a *Dendrobium* Orchid

Benjawan Lertwiriyawong[1], Kisana Phinija[2] and Pattana Srifah Huehne[1, 2]*

ABSTRACT

The *16S ribosomal RNA* (*16SrRNA*)-*ribosomal protein S12* (*rps*12) DNA located in the inverted repeat regions (IRs) of chloroplast DNA (cpDNA) can be used as an efficient targeting region for plastid vector construction in plastid transformation systems. Thus, the *16S rRNA-rps12* cpDNA sequence of *Dendrobium* Jacquelyn Thomas 'UH44-50' was cloned and analyzed. The 3,782 base pair (bp) *16SrRNA-rps12* DNA sequence isolated from this *Dendrobium* cultivar was 97% similar to the related *Phalaenopsis aphrodites* (AY916449) and *Oncidium* Gower Ramsey (GQ324949) cpDNA. It was composed of the highly conserved sequence of 1,427 bp *16SrRNA* at the 5′-end, 99 bp of *rps12* at the 3′-end and 72 bp of the transfer RNA Valine (GAC) (*trnV* $^{(GAC)}$). Nevertheless, the most variable regions were located in a 1,955 bp long spacer between the 3′-end *trnV* $^{(GAC)}$ and the 5′-end *rps12*. Apart from *Phalaenopsis* and *Oncidium* cpDNAs, the *16SrRNA-rps12* alignment of the *Dendrobium* nucleotide with nine other published sequences showed that this cpDNA region was unpredictably more homologous to *Phoenix dactylifera*, *Liriodendron tulipifera* and *Drimys granadensis* cpDNAs than those of grasses, including *Oryza japonica*, *Triticum aestivum* and *Zea mays*. The conserved and variable regions in the *16SrRNA-rps12* fragments of all twelve different species were located exactly in the same positions and orientation. Therefore, the highly conserved flanking regions of *16S rRNA-rps12 Dendrobium* cpDNA have potential to be an appropriate homologous sequence for plastid transformation in orchids.

Keywords: chloroplast DNA, inverted repeat, cpDNA spacer, *Dendrobium*, *16SrRNA*

INTRODUCTION

The plastid organelles generally contain circular double-stranded genomes 120-160 kilobases (kb) in size, having around 150 genes (Bock, 2001) with multiple copies (10,000 cpDNA copies or 100 chloroplast per cell and each chloroplast has 100 copies of its genetic material) in one cell (Bendich, 1987). Recently, only two complete orchid chloroplast genomes of *Phalaenopsis aphrodites* subsp. *formosana* (148 kb; Accession no. AY916449) and *Oncidium* Gower Ramsey (146 kb; Accession no.GQ324949) were available in the Genbank Database. Their encoding sequences and the structure of both orchid cp genomes were comparable in two 25 kb inverted repeats (IRa and IRb), an 85–86 kb large, single-copy region (LSC) and an 11–12 kb small, single-copy region (SSC) (Chang *et al.*, 2006; Wu *et al.,* 2010).

1 Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.

2 Laboratory of Biotechnology, Chulaborn Research Institute, Bangkok 10210, Thailand.

* Corresponding author, e-mail: fscipns@ku.ac.th

Received date : 15/11/10 Accepted date : 18/03/11

The plastid transformation has been described to eliminate many concerns of nuclear transformation (Bogorad, 2000; Maliga, 2002). This has been widely accepted because the plastid transformation is not only an environmentally friendly approach but it also results in containment of foreign genes and hyper expression. Due to the high copy number of plastids in one cell, the level of transgene expression and transcript accumulation was about 100-fold higher in transplastomic plants than in nuclear transgenic plants (Guda *et al*., 2000).

To achieve the plastid transformation, a cassette of transgenes in the plastid expression vector must consist of a chloroplast homologous sequence for the appliance to target the foreign gene into the precise location of the plastid genome. Thus, two chloroplast homologous sequences flanking *16SrRNA* and the transfer RNA Isoleucine (*trnI*) DNA in the IR region were first proposed by Svab *et al*. (1990) for stable plastid transformation in tobacco. Later on, Zoubenko *et al*. (1994) established the pPRV (plastid repeat vector) transformation vector with highly efficient targeting of the *trnV*$^{(GAC)}$*-rps12* region of tobacco chloroplast DNA. Another example for the pPRV vector family included pSBL-CTV2 targeting insertion in the *trnI-trnA* (transfer RNA Alanine) intergenic region and pRB vector targeting the *trnI-trnG* (transfer RNA Glycine) region (Maliga, 2002).

Due to the lack of knowledge of orchid plastid transformation, in the present study, the *16SrRNA-rps12* fragment from *Dendrobium* orchids was the first target region for cassette vector construction in plastid transformation technology. *Dendrobium* orchids belong to the Epidendroideae subfamily (Epidendroids). They are the most widespread orchids in Thailand. (Boonkorkaew, 2010). However, there are no *Dendrobium* cpDNA sequence data available for access in the database. The present study investigated the use of gene cloning and sequence analysis of the *16SrRNA-rps12* region of *Dendrobium* cpDNA, compared with *Phalaenopsis* and *Oncidium* orchid cpDNAs and nine other related cpDNA representatives from monocotyledonous and dicotyledonous plants, to clarify the conserved flanking region and genetic diversity based on nucleotide sequences in the *16SrRNA-rps12* region.

MATERIALS AND METHODS

DNA extraction

Total DNAs were extracted from samples of 200 mg of young leaves from *Dendrobium* Jacquelyn Thomas 'UH44-50' (*Dendrobium gouldii* × *Dendrobium phalaenopsis 'Lyon's Light No.1')* by the CTAB method (Murray and Thomson 1980). Briefly pre-warmed (60 °C) DNA extraction buffer (100 mM Tris-HCl, pH 8.0, 500 mM NaCl, 50 mM EDTA, 10 mM β-mercaptoethanol) was added to ground samples at a ratio of 1:2 (leaf tissue to buffer on a w/v basis), homogenized by vortex and incubated at 60 °C for 30 min. Then, 500 µL of chloroform:isoamyl alcohol mixture (24:1) was added and mixed for 30 sec. Starch and other cell debris were pelletted by 12,000×*g* centrifugation for 8 min at 4 °C and the supernatant was transferred to a new tube, to which 0.6 volume of 4M potassium acetate (pH 4.8) and twice the volume of absolute ethanol were added. The contents were mixed well by inversion and chilled on ice for 10 min, followed with precipitation by 12,000×*g* centrifugation for 8 min at 4 °C. The pellet was washed with 500 µL 70% ethanol and then 30 µL sterilized water was added to suspend the DNA.

Amplification of the *16SrRNA-rps12* inverted repeat region

The *16SrRNA-rps12* IRs regions of the cpDNA were amplified as two separate fragments, *16SrRNA-trnV* and *trvV-rps12*, from the total DNA extracted from the *Dendrobium* sample. The PCR

amplification of the 5′-*16SrRNA-trnV* fragment was performed with *i*-Taq™ DNA polymerase (iNtRON Biotechnology, Inc.) using 1529R primer 5′-GCG TGA AGA AGT GTC AAA CC-3′ (Rudnoy *et al*., 2002) and trnV144 primer 5′-AGG GAT ATA ACT CAG CGG TAG AG-3′. For the 3′-*16SrRNA-rps12* fragment, 1325-16S primer 5′-CTC AGC CTA CGG GGT ATT AGC A-3′ and rps12 primer 5′-GAT ATC TCA CAC CGG CGG GTA AAT CC-3′ were employed. The thermal cycling programs started with preheating at 94 °C for 1 min, followed by 35 cycles of 94 °C for 1 min, 60 °C for 1 min and 72 °C for 2.5 min and then a final extension at 72 °C for 15 min using a Mastercycler Gradient Eppendorf Thermocycler (USA). The PCR products were analyzed by 1% agarose gel electrophoresis, purified and then cloned into pGEM-T®(Promega).

Sequence analysis

DNA sequencing was performed on positive clones by Macrogen Inc. (Korea). To identify the sequence, DNA sequences were analyzed by the BLASTN 2.2.24 computer software (Stephen *et al*., 1997). The alignment was performed using the multiple alignment computer software, ClustalW2 (Thompson *et al*., 1994). Eleven *16SrRNA-rps12* sequences located in the IR region of the complete cpDNAs were randomly selected from two orchids, four monocotyledonous plants (consisting of one palm and three grasses) and five dicotyledonous plants (consisting of three shrubs and two perennial herbaceous plants) as shown in Table 1. The 11 sequences were extracted from the GenBank database and aligned with that of the *Dendrobium* sample to generate genetic relationship clusters. The phylogenetic tree was constructed by the neighbor-joining (NJ) method (Saitou and Nei, 1987) with 1,000 bootstrap replicates by the MEGA4 computer software (Tamura *et al*., 2007). The raw sequence data of the *Oncidium* and *Phalaenopsis* IR regions were extracted (Table 2) to obtain information on the length of the intergenic spacer (IGS).

Table 1 Twelve sources of *16SrDNA-rps12* cpDNAs used in this study.

Plant Species/common name	Native	Accession #	Size (bp)	Spacer (bp)
1. *Dendrobium* Jacquelyn 'UH44-50'/-	*Den. gouldii* x *Den. phalaenopsis*	HQ540308	3,782	28/1957
2. *Phalaenopsis Aphrodite*/Moth orchids	Taiwan	AY916449	3,672	36/1935
3. *Oncidium* Gower Ramsey/-	*O*. Goldiana x *O*. Guinea Gold	GQ324949	3,722	31/1890
4. *Phoenix dactylifera*/ true date palm	West Asia	GU811709.2	3,696	27/1908
5. *Liriodendron tulipifera*/ tulip poplar	North America	DQ899947.1	3,668	31/1921
6. *Drimys granadensis*/ *Tasmannia aromatica*	Australia	DQ887676.1	3,479	28/1965
7. *Nicotiana tabacum*/ Tobacco	America	Z00044.2	3,440	27/1952
8. *Ranunculus macranthus*/ buttercup	North America	DQ359689.1	3,693	27/1941
9. *Oryza japonica*/ Japanese rice	Japan	AY522330.1	3,559	27/1969
10. *Triticum aestivum*/ common wheat	Turkey	AB042240.3	3,555	30/1963
11. *Zea mays* /maize, corn	Mexico	X86563.2	3,593	31/1928
12. *Arabidopsis thaliana*/ thale cress	Europe, Asia	AP000423.1	3,752	27/1925

Table 2 Total intergenic spacer (IGS) regions in inverted repeat b (IRb) region of *Oncidium* and *Phalaenopsis* cpDNA.

Gene	Oncidium (GQ324949)		Phalaenopsis (AY916449)	
	Gene Position	Spacer (bp)	Gene Position	Spacer (bp)
rps19	82544- 82822	123	86240- 86518	124
*trn*H	82945- 83019	44	86642- 86715	44
rpl2	83063- 84548	19	86759- 88244	19
rpl23	84567- 84848	166	88263- 88544	166
*trn*I$^{(CAU)}$	85014- 85087	69	88710- 88783	69
ycf2	85156- 91821	1009	88852- 95724	958
trnL$^{(CAA)}$	92830- 92908	1119	96682- 96762	575
*ndh*B	94027- 96251	371	97337- 99365	333
rps12	99117-101007	1890	102851-104786	1935
*trn*V$^{(GAC)}$	99095- 99166	234	102858-102929	239
rrn16S	99400-100889	381	103168-104657	303
*trn*I$^{(GAU)}$	101270-102282	1381	104960-105975	1065
*trn*A$^{(UGC)}$	102269-103135	146	106039-106913	164
rrn23	103281-106087	99	107059-109865	99
rrn4.5	106186-106288	213	109964-110066	-
rrn5	106501-106621	235	-	-
*trn*R$^{(ACG)}$	106856-106929	-	110635-110708	-

ndh = NADH-plastoquinone oxidoreductase; *rpl* = ribosomal protein L; *rps* = ribosomal protein S; *rrn* = ribosomal RNA; *trn* = transfer RNA; *ycf* = conserved hypothetical protein ycf2

RESULTS

16SrRNA-rps12 gene cloning

Two PCR products, 1.8 and 2.5 kb, were obtained from the amplification of the *16SrRNA-rps12* cpDNA isolated from the *Dendrobium* sample. The *16SrRNA-trnV* 1.8 kb fragment, the 5′-end of the target DNA, was generated by 1529R and trnV144 primers and cloned into pDJT16S1004. The second amplified DNA fragment of the *trnV*$^{(GAC)}$*-rps12* spacers region (2.5 kb) was generated by the 1325-16S and rps12 primers. After gene cloning, a positive clone was obtained as pDJTrps8 (Figure 1).

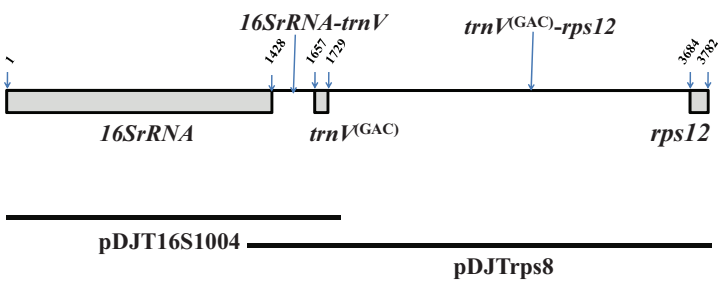

Figure 1 Map of overlapping *16SrRNA-trnV* (pDJT*16S*1004) and *trnV*$^{(GAC)}$ *-rps12* (pDJT*rps*8) cpDNA fragments that were used to determine the complete sequence of *Dendrobium 16SrRNA-rps12* cpDNA which consisted of three coding sequences of *16SrRNA, trnV*$^{(GAC)}$ and *rps12* and two noncoding sequences of *16SrRN-trnV*$^{(GAC)}$ and the *trnV*$^{(GAC)}$ *-rpn12* spacer.

The assembly of both overlapping sequences in pDJT16S1004 and pDJTrps8 was 3,782 bp in length, consisting of 1,427 bp partial *16SrRNA*, 229 bp *16SrRNA-trnV*$^{(GAC)}$ spacer and 72 bp *trnV*$^{(GAC)}$, 1,955 bp *trnV*$^{(GAC)}$*-rps12* spacer and 99 bp partial *rps12* of the inverted repeat region (IR) chloroplast genome (Figure 2). Nevertheless, the IR regions of these orchids contained the longest IGS DNA fragment, *trnV*$^{(GAC)}$*-rps12* spacer, with a range of 1890 - 1,955 bp in length (Table 2). The total 3,782 bp *16SrRNA-trnV*$^{(GAC)}$*-rps12* cpDNA isolated from the *Dendrobium* sample was deposited in the GenBank database under the accession number HQ540308.

Sequence comparison of the 16*SrRNA-rps12* IR region of cpDNAs among three orchid genera

Comparing the nucleotide sequences of the *Dendrobium 16SrRNA-rps12* IR region with those of published complete cpDNA of *Phalaenopsis* and *Oncidium* samples revealed 97% nucleotide identity (without gap) homologous to the nucleotide in the *16SrRNA-rps12* IR region of the *Oncidium* and *Phalaenopsis* cpDNAs. The nucleotide alignment of the *Dendrobium, Oncidium* and *Phalaenopsis 16SrRNA-rps12* cpDNAs showed clearly that both the 5′-end 1,427 bp *16SrRNA* and the 3′-end 99 bp *rps12* of the *16SrRNA-rps12* were highly conserved sequences and identical. On the other hand, the most variable regions, with 90–96% similarity to the *Oncidium* and *Phalaenopsis* cpDNAs, were located in a 1,955 bp *trnV*$^{(GAC)}$ *-rps12* spacer between the 1729 and 3683 nucleotide position of the *Dendrobium* cpDNA (Table 3). Five large gaps of 19–81 bp deletion (81, 41, 20, 19 and 42 bp) were disclosed in this spacer among the three orchid cpDNAs and a gap as small as 4–5 bp was placed in the *16SrRN-trnV*$^{(GAC)}$ spacer (1428 to 1656 nucleotide position) as shown in Figure 3.

Phylogenetic analysis

The genetic relationships between the nucleotide sequence of *Dendrobium* 16*SrRNA-rps12* and the other eleven 16*SrRNA-rps12* cpDNAs available in the GenBank database are listed in Table 1 and indicated that, beside the orchids, the *Dendrobium* sequence had 95, 94 and 94% nucleotides identical to those of *Phoenix dactylifera* (true date palm) and *Liriodendron tulipifera* and *Drimys granadensis*, (ornamental shrubs), respectively (Figure 4). The monocotyledonous grass plants, *Oryza japonica*, *Triticum aestivum* and *Zea mays*, and the dicotyledonous plants, consisting of *Nicotiana tabacum, Ranunculus macranthus* and *Arabidopsis thaliana,* had less sequences identical, with 91.5% on average. The topology of the phylogenetic tree was well supported, with three main group divisions. The members of each group were the five dicotyledonous plants, except for the date palm, three orchids and the three grasses with the exclusions having bootstrap values of 91, 100 and 100%, respectively (Figure 4). However, the bootstrap value of the subclade constructed between *Dendrobium-Phalaenopsis* and *Triticum-Oryza* was slightly lower at 49% when phylogenetic analyses in the gap sites were excluded and 1,000 repeat cycles for bootstrapping analyses were performed.

Based on an overall observation of the alignments of the total twelve 16*SrRNA-rps12* cpDNAs, the pattern of nucleotide-conserved regions performed in Figure 3 was the same for all three orchids. Only sequences in the 1,955 bp *trnV*$^{(GAC)}$*-rps12* spacer were highly divergent, with several gaps (Table 3 and Figure 3). The most variable region was located exactly in the same region as the 1,955 bp *trnV*$^{(GAC)}$*-rps12* spacers between the 1729 to 3683 nucleotide position of the *Dendrobium* cpDNA, but differed in size and had larger numbers. Apart from the grass group, the two *trnV*$^{(GAC)}$*-rps12* spacers of *N. tabacum* and *D. granadensis* presented the largest gap (greater than 200 bp) at different positions. The cluster patterns constructed using the two spacer nucleotide sequences were slightly different in

```
   1 tagccctgcc ttcggcatcc ccctccttgc ggttaaggta acgacttcgg gcatggccag
  61 ctcccatagt gtgacgggcg gtgtgtacaa ggcccgggaa cgaattcacc gccgtatggc
 121 tgaccggcga ttactagcga ttccggcttc atgcaggcga gttgcagcct gcaatccgaa
 181 ctgaggacgg gttttggagt tagctcaccc tcgcgggatc gcgacccttt gtcccggcca
 241 ttgtagcacg tgtgtcgccc agggcataag gggcatgatg acttgacgtc atcctcacct
 301 tcctccggct tatcaccggc agtctgttca gggttccaaa ctcaatggtg gcaactaaac
 361 acgagggttg cgctcgttgc gggacttaac ccaacacctt acggcacgag ctgacgacag
 421 ccatgcacca cctgtgtccg cgttcccgaa ggcacccctc tctttcaaga ggattcacgg
 481 catgtcaagc cctggtaagg ttcttcgctt tgcatcgaat taaaccacat gctccaccgc
 541 ttgtgcgggc ccccgtcaat tcctttgagt ttcattcttg cgaacgtact ccccaggcgg
 601 gatacttaac gcgttagcta cagcactgca cgggtcgata cgcacagcgc ctagtatcca
 661 tcgtttacgg ctaggactac tggggtatct aatcccattt gctcccctag ctttcgtctc   1..1427
 721 tcagtgtcag tgtcggccca gcagagtgct ttcgccgttg gtgttctttc cgatctctac    <16SrRNA>
 781 gcatttcacc gctccaccgg aaattccctc tgcccctacc gtactccagc ttggtagttt
 841 ccaccgcctg tccagggttg agccctggga tttgacggcg gacttgaaaa gccacctaca
 901 gacgctttac gcccaatcat tccggataac gcttgcatcc tctgtcttac cgcggctgct
 961 ggcacagagt tagccgatgc ttattcctca gataccgtca ttgcttcttc tccgagaaaa
1021 gaagttcacg acccgtgggc cttctacctc cacgcggcat tgctccgtca ggctttcgcc
1081 cattgcggaa aattccccac tgctgcctcc cgtaggagtc tgggccgtgt ctcagtccca
1141 gtgtggctga tcatcttctc ggaccagcta ctgatcatcg ccttggtaag ctattgcctc
1201 accaactagc taatcagacg cgagcccctc ctcgggcgga ttcctccttt tgctcctcag
1261 cctacggggt attagcaacc gtttccagtt gttgttcccc tcccaagggc aggttcttac
1321 gcgttactca cccgtccgcc actggaaaca ccacttcccg tccgacttgc atgtgttaag
1381 catgccgcca gcgttcatcc tgagccagga tcgaactctc catgagattc atagttgcat   1428..1656
1441 tacttatagc ttccttgttc gtagacaaag ctgattcgga attgtctttc attccaaggc    <16SrRNA-trnV
1501 cgaaggcttg tatccatgcg cttcatatta gcctggagtt cgctcccagc aatatagcca    Spacer>
1561 tccctaccct atcacgtcaa tcccacaagc ctcttatcca ttctcgttcg atcgcggcgg
1621 gggagcaagt ccaaatagaa aaactcccat tgggttagg gataatcagg ctcgaactga   1657..1728
1681 tgacttccac cacgtcaagg tgacactcta ccgctgagtt atatccctc ccggtcccat    <trnV(GAC)>
1741 cgagaaatag aactaactaa tcctaaggca aagggtcgag aaactaaacg ccactattcc
1801 tgaacaactt ggagccgggc cttcttttcg cactattatg ttatggatag tcaaataatg
1861 ggaaaaattg gattccattg tcaactgctc ctatcggaaa taggattgac tacggattcg
1921 agccatagca catggattcc taaaatccgt tttcccgatc taaatcgaac agtttttaca
1981 tgaagaagat ttttttcagc atgctctatt cgaggctggt aggagaagaa ccgactcggt
2041 attgtaaaaa aaaaagagtg gaagcagaac caagtcaaga tgatatgaat caccccttct
2101 tcttgcgcca aggatcttac taccatttcc gaaggaactg gagctacttt tcttttccat
2161 ttccattcca gagttcctat gtgtttccac gccctttct taggaacaca tacaataaaa
2221 aaggataacg gtaaccccac cattaactac ttaattttca tttatgaatt gaatagtaat
2281 agaaatacat gtcctactga gacagagaca gaatttgtaa cttgctatcc tcttgcctag
2341 caggcaaaga tggacccccg tggagagact gattcattcg gatcgacatg agggtccaac
2401 tccattgcat tgccagaatc catgttgtat atttgaagca gttgacctcc ttgcttctct
2461 catggtacaa ccctcttccc gcggagcccc ttttttcctc ggtccacaga gacaaaatat
2521 aggactggtg ccaacagttc atcacggaag aaaggactcg ccgagccggg atcgctaact   1729..3683
2581 aatactaaaa ctaagactat aagactaata taactaatct aattagaaat aattctaata    <trnV-rns12
2641 attagataat tagaatatct aagaattaga atataataat aatagaataa tacataatag    Spacer>
2701 aaaagatata atagaaaaga actgtctttt ctgtatactt tccccggttc cgttgctacc
2761 gcgcccttac gcaatcgatc ggatcatata gatttcaaca caacataggt catcgaaagg
2821 atctcggacg actcaccaaa gcacgaaagc caggatcttt cagaaaatgg attcctattt
```

```
2881 gaagagtgca taaccgcatg gataagctca cactaacccg tccattttgg atccaattcg
2941 ggattttcct taggaagtat cgggaaggaa ttggaatgta agaatatcga ttcatacaga
3001 agaaaaggtt ctctattgat tcaaacgcta tacctatggg atagaggagg aggaaaaccc
3061 gaagatttca catagtactt ttgatcaaaa aagaaatctg atttctttcg tacccctcgc
3121 tcgatgagaa aatgggtcag attctacagg atcaacaaac ctatgggact tcaggaatga
3181 tggaagggaa tcgaaataaa atcgaaatga aagaaaaatc caaaagaaag aagaaaacaa
3241 aaaaaaaaag aagtaaatct aaaaagaaat agaaaagaaa agtagaagat agaagagccc
3301 agattccaga tgaacaaatt caaactcgaa aaggatattt ctggttctcg aagaatgagg
3361 ggcaaaggga ttgatcgaga gaagatctct tgttcttctt ataattagaa gatcgtgatt
3421 tgctccgcat aggtttggta aaaagaagaa tcttctcctt tgatcataaa tggaaagtgt
3481 tcaattggaa catgaaaacg tgaatgaatt ggtctgagtt actcttcggg acggagtaga
3541 agaagggagg agattctcga acgaagaaaa ggatccaatt acttcgaaag aattgaacga
3601 ggagccgtat gaggtgaaaa tctcatgtac ggttctgtag agtgacagta agggtgactt
3661 atctgtcaac ttttccacta tcacccccaa aaaaccaaac tctgccttac gtaaagttgc   3684..3782
                                                                      <rps12>
3721 cagagtacga ttaacctctg gatttgaaat cactgcttat atacctggta ttggccataa
3781 tt
```

Figure 2 Nucleotide sequence of the 3,782 bp *16SrRNA-rps12* inverted repeat region of *Dendrobium* chloroplast DNA. This IR region encoded a partial *16S rRNA* (1..1490) at the 5′-end, complete *trnV*^{GAC} (1657..1728) and a partial *rps12* (3684..3782) at the 3′-end. There are two spacers, *16SrRNA-trnV* spacer (1428..1656) and *trnV-rns12* spacer (1729..3683) that were located in between each gene. All genes are indicated as bold letters.

Table 3 Percent similarity of genes and intergenic spacer (IGS) in *16S-rps12* cpDNA regions resulting from *Dendrobium, Oncidium* and *Phalaenopsis* cpDNA comparison.

Pair-wise	Gene and IGS/size (bp)					
	16S-rps12 (3782)	*16S** (1427)	*16S*-IGS (195)	*trnV* (106)	*trnV*-IGS (1955)	*rps12** (99)
Dendrobium/*Oncidium*	97	99	97	100	92	100
Dendrobium/*Phalaenopsis*	97	99	98	100	90	100
Oncidium/*Phalaenopsis*	95	99	98	100	96	100

*= Partial gene sequence.

topology (Figure 4). The phylogenetic tree based on the 1,955 bp *trnV*$^{(GAC)}$*-rps12* spacers defined the genetic relationship between members of each cluster better than the phylogenetic tree created by the 229 bp *16SrRNA-trnV*$^{(GAC)}$ spacer sequences and was supported well by the whole length of the 3,782 bp 16*SrRNA-rps12* tree with better confidence values for the bootstrap values of over 70%.

DISCUSSION

In the present study, the sequence conservation and diversity in the IR region at the *16SrRNA-rpn12* position were investigated by comparing the nucleotide sequences of the cloned *16SrRNA-rpn12 Dendrobium* cpDNA with the same cpDNA region of *Phalaenopsis*, *Oncidium* and nine other closely related sequences extracted from the GenBank database. The *16SrRNA-rpn12* cpDNA isolated from the *Dendrobium* sample was

(A)

***16SrRNA-trnV*$^{(GAC)}$ Spacer** (229 bp)

Small gap (4-5 bp)

```
Dendrobium      CGAAGGCTTGTATCCAT----GCGCTTCATATTAGCCTGGAGTTCGCTCCCAGCAATATA 1556
Phalaenopsis    CGAAGGCTTGTATCCATCCATGCGCTTCATATTAGCCTGGAGTTCGCTCCCAGCAATATA 1560
Oncidium        CGAAGGCTTGTATCCATCCATGCGCTTTATATTAGCCTGGAGTTCGCTCCCAGCAATATA 1560
                *****************    ****** ******************************

Dendrobium      GCCATCCCTACCCTATCACGTCAATCCCACAAGCCTCTT-----ATCCATTCTCGTTCGA 1611
Phalaenopsis    GCCATCCCTACCCTATCACGTCAATCCCACAAGCCTCTTCTCTTATCCATTCTCGTTCGA 1620
Oncidium        GCCATCCCTACCCTATCACGTCAATCCCACAAGCCTCTT-----ATCCATTCTCGTTCGA 1615
                ************************************     ****************
```

(B)

***trnV*$^{(GAC)}$*-rps12* Spacer** (1,955 bp)

Gap 1 (81 bp)

```
Dendrobium      CACTATTCCTGAACAACTTGGAGCCGGGCCTTCTTTTCGCACTATTATGTTATGGATAGT 1851
Phalaenopsis    CACTATTTCTGAACAACCTGGGGCCGGGCCTTCTTTTCGCACTATTATGTTATGGATAGC 1860
Oncidium        CACTATT------------------------------------------------------ 1802
                *******

Dendrobium      CAAATAATGGGAAAAATTGGATTCCATTGTCAACTGCTCCTATCGGAAATAGGATTGACT 1911
Phalaenopsis    CAAATAATGGGAAAAATTGGATTCCATTGTCAACTGCTCCTATCGGAAATAGGATTGACT 1920
Oncidium        ----------------------------GTCAACTGCTCCTATCGGAAACAGGATTGACT 1834
                                            ******************** **********
```

Gap 2 (41 bp)

```
Dendrobium      GGAACACATACAATAAAA------------------------AAGGATAACGGTAACCCC 2238
Phalaenopsis    G----------------------------------------AAGGATAACG-----CCC 2225
Oncidium        GGAACACATACAATAAAAATCTTAGGAACACATACAATAAAAAAGGATAACG-----CCC 2189
                *                                        **********     ***
```

Gap 3 (20 bp)

```
Dendrobium      ACTAAGACTATAAGACTAATATAACTAATCTAATTAGAAATAATTCTAATAATTAGATAA 2649
Phalaenopsis    ACTAAGACTATAAGACTAATATAACTAATCTAATTAGAATT------------------- 2613
Oncidium        ACTAAGACTATAAGACTAATAGAACTAATCTAATTAGAAATAATTCGAATAAT------- 2593
                ******************** ***************** *
```

Gap 4 (19 bp)

```
Dendrobium      TAATAGAAAAGAACTGTCTTTTCTGTATACTTTCCCCGGTTCCGTTGCTACCGCG-CCCT 2767
Phalaenopsis    GAATAGAAAAGAACTGTCTTTTCTGT-------------------TGCTACCGCGGCCCT 2694
Oncidium        TAATAGAAAAGAACTGTCTTTTCTTT-------CCCCGGTTCCGTTGCTACCGCGGCCCT 2696
                 ********************** *                  ********** ****
```

Gap 5 (42 bp)

```
Dendrobium      AAAAAAAAA--AGAAGTAAATCTAAAAAGAAATAGAAAAGAAAAGTAGAAGATAGAAGAG 3297
Phalaenopsis    ------------------------------------AAAAGTAGAAGATAGAAGAG 3188
Oncidium        AAAAAGAAAGAAGAAGTAAATCTAAAAATAAATCTAAAAAAAAAGTAGAAGATAGAAGAG 3236
                                                    ********************
```

Figure 3 Nucleotide alignments of the most variable regions located in the 16*SrRNA-rps12* IR cpDNAs of the three orchids, *Dendrobium* Jacquelyn 'UH44-50' (HQ540308), *Phalaenopsis aphrodite* subsp. *formosana* (AY916449) and *Oncidium* Gower Ramsey (GQ324949). (A) 229 bp *16SrRNA-trnV*$^{(GAC)}$ spacer; and (B) 1,955 bp *trnV*$^{(GAC)}$*-rps12* spacer. * = identical nucleotides represented and - = absence of nucleotide (gap).

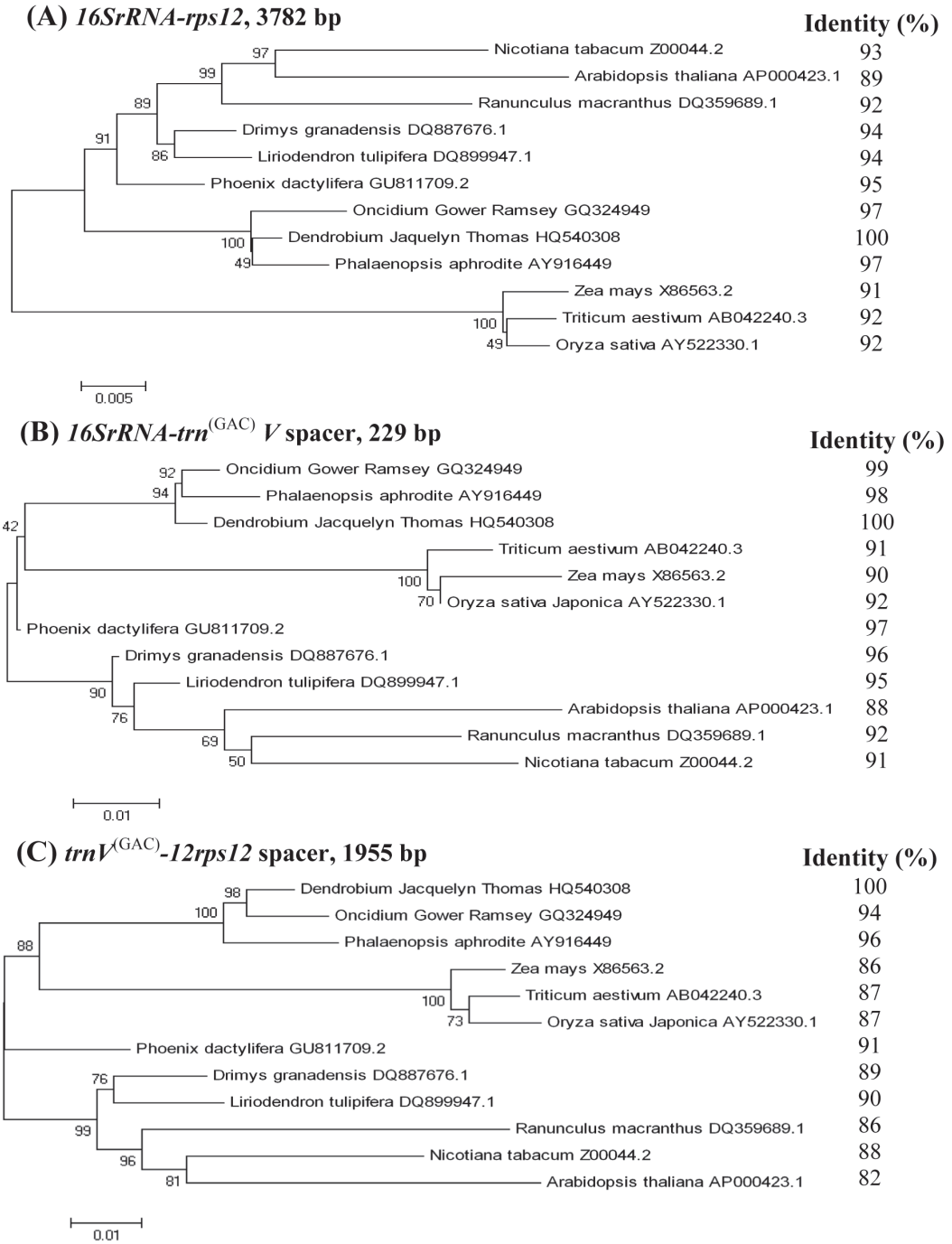

Figure 4 Neighbor-joining phylogenetic analysis based on the nucleotide variation of the 16*SrRNA-rps12* IR region in the chloroplast genome of *Dendrobium* compared with the eleven reported complete cp genomes in GenBank: (A) 3,782 bp *16SrRNA-trnV* phylogenetic relationship; (B) 229 bp *16SrRNA-trnV*$^{(GAC)}$ spacer phylogenetic tree; and (C) 1,955bp *trnV*$^{(GAC)}$*-rps12* spacer phylogenetic tree. The clustering was performed with 1,000 replicates for bootstrapping analysis using the MEGA4 computer software, excluding gap scoring.

found to be the longest fragment (3,782 bp). It had almost the same length as that of the *Arabidopsis* cpDNA fragment, while the lengths of the *Phalaenopsis* and *Oncidium* cpDNAs were 3,672 and 3,722 bp, respectively. In contrast, the cpDNAs of tobacco and the three grasses (rice, maize and wheat) were much shorter than that of the *Dendrobium* sample (Table 1). This result was supported by Chang *et al.* (2006) who reported that the IR sequences of the grasses are much shorter than those of *Phalaenopsis* sp. and tobacco.

The cloned *16SrRNA-rpn12 Dendrobium* cpDNA covered three conserved coding sequences of partial *16SrRNA* (5′-flanking region), *trnV*$^{(GAC)}$ and partial *rps12* (3′-flanking region), and two variable noncoding sequences of *16SrRN-trnV*$^{(GAC)}$ and the *trnV*$^{(GAC)}$*-rpn12* spacer. The overall *Dendrobium*, *Oncidium* and *Phalaenopsis* features in gene organization, position and orientation had the same pattern along the *16SrRNA-rpn12* regions. By comparing total intergenic spacers in the IR region with raw sequences between *Oncidium* and *Phalaenopsis* cpDNA, the *trnV*$^{(GAC)}$*-rps12* spacer was found to be the longest. The difference in size between these cpDNA regions occurred mainly in the 1,955 *trnV*$^{(GAC)}$*-rpn12 Dendrobium* spacer that correlated with a number of nucleotide insertions and deletions in this spacer.

Several large gaps in this spacer were concurrent with the nucleotide variation in the spacers of all *trnV*$^{(GAC)}$*-rpn12* cpDNAs from the twelve plant genera used in the present study (Table 3). *Dendrobium* (Subfamily Epidendroideae / Tribe Epidendreae / Subtribe Dendrobiinae) had a growth habit similar to that of *Oncidium* (Epidendroideae / Cymbidieae / Oncidiinae) to a greater extent than to *Phalaenopsis* (Epidendroideae / Vandeae / Aeridinae. The sequence of *Dendrobium 16S-rps12* (either the full-length or the *trnV*$^{(GAC)}$*-rps12* fragment) was more closely related to *Oncidium* than to *Phalaenopsis*. This may have been due to the fact that both of the genera *Dendrobium* and *Oncidium* are of sympodial growth while the genus *Phalaenopsis* has monopodial growth. The reed-stem habit is thought to be ancestral within the epidendroids (Dressler, 1993).

Comparing the phylogenetic tree structures generated by full-length sequences (3,782 bp), the *16SrRNA-trnV*$^{(GAC)}$ spacer (299 bp) and the *trnV*$^{(GAC)}$*-rpn12* spacer (1,955 bp) of the twelve cpDNAs, the tree could be best classified into three main clusters (orchids, grasses and eudicots), related to their genetic relationships by the NJ method (Figure 4). The highly conserved sequences at the 5′end (*16SrRNA* region) and the 3′end (*rps12* region) of the flanking region among the three orchid genera (*Dendrobium, Oncidium* and *Phalaenopsis*) and the other nine representatives of monocotyledonous and dicotyledonous cpDNA were evidently established at both ends of all fragments. Therefore, the use of the *16SrRNA-rps12* fragment of *Dendrobium* cpDNA as the homologous sequences for the plastid vector cassette showed great potential for orchid plastid transformation. Many reports have shown that the tobacco-specific plastid vector could be achieved for potato (Sikdar *et al.,* 1998) and tomato (Ruf *et al.,* 2001); therefore it is not necessary to construct a specific plastid transformation vector (Maliga, 2002).

In addition, the 1,955 *trnV*$^{(GAC)}$*-rpn12* spacer phylogenetic tree topology presented the best clusters. Each cluster of the *trnV*$^{(GAC)}$*-rpn12* spacer phylogenetic tree was stable with high bootstrapping values (greater than 70%) and were simply distinguished by the percentage of nucleotide identity. Moreover, the Orchidaceous members (*Dendrobium, Oncidium and Phalaenopsis*) and the Gramineae or Poaceae grasses (rice, maize and corn) were split into more closely related clusters within the monocotyledonous plants (including the true date palm, *Phoenix dactylifera*) and some trees (the tulip poplar and *Tasmannia aromatica*), rather than

with herbaceous plants, such as tobacco and *Arabidopsis* sp. (Figure 4C). The present study supported the *Amborella*-*Nymphaea* / *Amborella*-basal hypotheses proposed by Goremykin *et al*. (2003) and Soltis and Soltis (2004) that the *Amborella* dicotyledonous plant is not a sister to the remaining flowering plants, and the *Oncidium* monocotyledonous plant could occupy the *Amborella* position in the phygenetic tree. The root of angiosperm phylogeny was at the basal monocot-dicot split and that the grasses (Poaceae) were the deepest branch in the phylogeny of the angiosperms (Goremykin *et al.,* 2003). However, the gene and intron content are highly conserved among the early diverging angiosperms and basal eudicotyledons, and lineage-specific correlation was detected between rates of nucleotide substitutions and genomic rearrangements (Jansen *et al*., 2007).

In the present report, the conserved genetic variation based on base deletion and insertion was demonstrated to be found commonly in the *trnV*$^{(GAC)}$-*rpn12* spacer of both monocotyledonous and dicotyledonous plants. The appropriately designed nucleotide sequence as a barcode (300–800 bp) in this spacer could be a high potential chloroplast marker if it is easy to perform and can resolve higher-level relationships with high levels of confidence at the intra- and inter-specific level. Several noncoding cpDNA fragments have been established for phylogenetic analysis and genetic identification as plant DNA barcoding, such as the transfer RNA *Leucine* (*trnL*) intron and the transfer RNA *Leucine*-*Phynylalanine* (*trnL-F*) spacer (Taberlet *et al*., 1991), and the transfer RNA *Lysine* (*trnK*) introns (Johnson and Soltis, 1994; van den Berg *et. al*., 2009). However, the molecular genetic classifications in some groups of orchids are still limited and uncertain. Further studies are required to determine the genetic relationships in various plant species if the gaps are to be included for multiple alignment and bootstrapping analyses.

CONCLUSION

The highly conserved region at both ends of cloned *16S rRNA-rps12 Dendrobium* orchid cpDNA will be advantageous as a source of informative data for orchid plastid vector construction. Meanwhile, the orchid cpDNA spacer between *trnV* $^{(GAC)}$ and *rps12,* containing the most variation in nucleotide sequencing, has great potential as an excellent effective cpDNA marker for complex orchid classification.

ACKNOWLEDGEMENTS

This study was supported by the Golden Jubilee Program, Thai Research Fund (TRF) contract no. BGJ458009, Chulabhorn Research Institute (CRI) and the Center for Advanced Studies in Tropical Natural Resources, Kasetsart University, Thailand.

LTERATURE CITED

Bendich A.J. 1987. Why do chloroplasts and mitochondria contain so many copies of their genome? **BioEssays** 6: 279–282.

Bock, R. 2001. Transgenic plastids in basic research and plant biotechnology. **J. Mol. Biol.** 312: 425–438.

Bogorad, L. 2000. Engineering chloroplasts: An alternative site for foreign genes, proteins, reactions and products. **Trends Biotechnol.** 18: 257–263.

Boonkorkaew, P. 2010. **List of Biodiversity Resource: Thai Orchid Database.** Biodiversity Resource Inventory Development Project, Biodiversity-based Economics Development Office (Public Organization), Bangkok.

Chang, C., I. Pin, T. Chow, H. Chen, W. Chen, C. Cheng, C. Lin, S. Liu, C. Chang and S. Chaw. 2006. The chloroplast genome of *Phalaenopsis aphrodites* (Orchidaceae): Comparative

analysis of evolutionary rate with that of grasses and its phylogenetic implications. **Mol. Biol. Evol.** 23: 279–291.

Dressler, R.L. 1993. **Phylogeny and Classification of the Orchid Family.** Cambridge University Press. Port Melbourne, Australia. 314 pp.

Jansen, R.K., Z. Cai, L.A. Raubeson, H. Daniell, C.W. dePamphilis, J. Leebens-Mack, K.F. Müller, M. Guisinger-Bellian, R.C. Haberle, A.K. Hansen, T.W. Chumley, S. Lee, R. Peery, J.R. McNeal, J.V. Kuehl and J.L. Boore. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. **Proc. Natl. Acad. Sci. USA** 104: 19369–19374.

Johnson, L.A. and D.E. Soltis. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae *sensu stricto*. **Syst. Bot.** 19: 143–156.

Goremykin, V.V., K.I. Hirsch-Ernst, S. Wölfl and F.H. Hellwig. 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. **Mol. Biol. Evol.** 20: 1499–1505.

Guda, C., S.B. Lee and H. Daniell. 2000. Stable expression of a biodegradable protein-based polymer in tobacco chloroplasts. **Plant Cell Rep.** 19: 257–262.

Maliga, P. 2002. Engineering the plastid genome of higher plants. **Curr. Opin. Plant Biol.** 5: 164–172.

Murray, M.G. and W.F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. **Nucleic Acids Res.** 8: 4321–4326.

Ruf, S., M. Hermann, I.J. Berger, H. Carrer and R. Bock. 2001. Stable genetic transfomation of tomato plastids and expression of a foreign protein in fruit. **Nat. Biotechnol.** 19: 870–875.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Mol. Biol. Evol.** 4: 406–425.

Sikdar, S.R., G. Serino, S. Chaudhuri and P. Maliga. 1998. Plastid transformation in *Arabidopsis thaliana***. Plant Cell Reptr.** 18: 20–24.

Soltis, D.E. and P.S. Soltis. 2004. *Amborella* not a "basal angiosperm"? Not so fast. **AM. J. Bot.** 91: 997–1001.

Stephen, F.A., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. **Nucleic Acids Res.** 25: 3389–3402.

Svab Z., P. Hajdukiewicz and P. Maliga. 1990. Stable transformation of plastid in higher plants. **Proc. Natl. Acad. Sci. USA** 90: 913–917.

Taberlet, P., L. Gielly, G. Patou and J. Bouvet. 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. **Pl. Mol. Biol.** 17: 1105–1109.

Tamura, K., J. Dudley, M. Nei and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Mol. Biol. Evol.** 24: 1596–1599.

Thompson, J.D., D.G. Higgins and T.J. Gibson. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. **Nucleic Acids Res.** 22: 4673–4680.

van den Berg, C., W.E. Higgins, R.L. Dressler, W.M. Whitten, M.A. Soto-Arenas and M.W. Chase. 2009. A phylogenetic study of Laeliinae (Orchidaceae) based on combined nuclear and plastid DNA sequences. **Ann. Bot.** 104: 417–430.

Wu, F.H., M.T. Chan, D.C. Liao, C.T. Hsu, Y.W. Lee, H. Daniell, M.R. Duvall and C.S. Lin. 2010. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. **BMC Plant Biol.** 10: 68 (doi: 10.1186/1471-2229-10-68).

Zoubenko, O.V., L.A. Allison, Z. Svab and P. Maliga. 1994. Efficient targeting of foreign genes into the tobacco plastid genome. **Nucleic Acids Res.** 22: 3819–3824.