Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae

1-1-2010

Fu-Hui Wu
Ming-Tsair Chan
De-Chih Liao
Chen-Tran Hsu
Yi-Wei Lee

See next page for additional authors

Find similar works at: https://stars.library.ucf.edu/facultybib2010

University of Central Florida Libraries http://library.ucf.edu

Recommended Citation

Wu, Fu-Hui; Chan, Ming-Tsair; Liao, De-Chih; Hsu, Chen-Tran; Lee, Yi-Wei; Daniell, Henry; Duvall, Melvin R.; and Lin, Choun-Sea, "Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae" (2010). Faculty Bibliography 2010s. 944.
https://stars.library.ucf.edu/facultybib2010/944

This Article is brought to you for free and open access by the Faculty Bibliography at STARS. It has been accepted for inclusion in Faculty Bibliography 2010s by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae

Fu-Hui Wu†1, Ming-Tsair Chan†1, De-Chih Liao1, Chen-Tran Hsu1, Yi-Wei Lee1, Henry Daniell2, Melvin R Duvall3 and Choun-Sea Lin*1

**Abstract**

**Background:** *Oncidium* spp. produce commercially important orchid cut flowers. However, they are amenable to intergeneric and inter-specific crossing making phylogenetic identification very difficult. Molecular markers derived from the chloroplast genome can provide useful tools for phylogenetic resolution.

**Results:** The complete chloroplast genome of the economically important *Oncidium* variety *Onc.* Gower Ramsey (Accession no. GQ324949) was determined using a polymerase chain reaction (PCR) and Sanger based ABI sequencing. The length of the *Oncidium* chloroplast genome is 146,484 bp. Genome structure, gene order and orientation are similar to *Phalaenopsis*, but differ from typical Poaceae, other monocots for which there are several published chloroplast (cp) genome. The *Onc.* Gower Ramsey chloroplast-encoded NADH dehydrogenase (*ndh*) genes, except *ndhE*, lack apparent functions. Deletion and other types of mutations were also found in the *ndh* genes of 15 other economically important Oncidiinae varieties, except *ndhE* in some species. The positions of some species in the evolution and taxonomy of Oncidiinae are difficult to identify. To identify the relationships between the 15 Oncidiinae hybrids, eight regions of the *Onc.* Gower Ramsey chloroplast genome were amplified by PCR for phylogenetic analysis. A total of 7042 bp derived from the eight regions could identify the relationships at the species level, which were supported by high bootstrap values. One particular 1846 bp region, derived from two PCR products (trnH-GUG-psbA and trnF-GAA-ndhJ) was adequate for correct phylogenetic placement of 13 of the 15 varieties (with the exception of *Degarmoara* Flying High and *Odontoglossum* Violetta von Holm). Thus the chloroplast genome provides a useful molecular marker for species identifications.

**Conclusion:** In this report, we used *Phalaenopsis. aphrodite* as a prototype for primer design to complete the *Onc.* Gower Ramsey genome sequence. Gene annotation showed that most of the *ndh* genes in Oncidiinae, with the exception of *ndhE*, are non-functional. This phenomenon was observed in all of the Oncidiinae species tested. The genes and chloroplast DNA regions that would be the most useful for phylogenetic analysis were determined to be the *trnH-GUG-psbA* and the *trnF-GAA-ndhJ* regions. We conclude that complete chloroplast genome information is useful for plant phylogenetic and evolutionary studies in *Oncidium* with applications for breeding and variety identification.

**Background**

The Oncidiinae subtribe of the Orchidaceae family, consisting of about 70 closely related genera with over 1000 species, is divided into five alliances, with *Oncidium* as its largest genus [1]. From the perspective of cellular biology, ecology and morphology, *Oncidium* is the most diverse genus in the Orchidaceae. Traditionally, the taxonomy of the Oncidiinae tribe is based on the morphology of the flower [2]; however, morphology is affected by environmental factors, and over time flower morphologies have evolved convergently. The positions of some species in the evolution and taxonomy of Oncidiinae are therefore difficult to identify. Accurate identification is further complicated by the ease with which Oncidiinae can be crossed intergenerically, as indicated by the 107 interge-
neric hybrids reported [1] and the fact that more than 2200 hybrids (about 20% in the Oncidium group) have been re-distributed into other genera.

Different molecular marker techniques such as terminal restriction fragment length polymorphism (TRFL), arbitrarily primed polymerase chain reaction (AP-PCR), DNA amplification fingerprinting (DAF), and random amplification polymorphism DNA (RAPD) are available to conduct genetic analyses by PCR and provide information about evolution that is useful for taxonomy. Tsai et al. [3] used 257 RAPD markers to investigate the relationships between 24 species of Oncidiinae, and found that the species could be separated into seven groups; however, Tsai and colleagues were unable to identify the more detailed relationships among these species.

Although there are three different genomes in plants, chloroplast DNA (cpDNA) is in many respects the genome of choice for taxonomic studies in orchids [2] as well as other species [4,5]. There are many advantages to using cpDNA for taxonomy and evolutionary research: (1) the size of cpDNA is small, with high copy number and simple structure; (2) when compared to the mitochondrial and nuclear genome, cpDNA gene content and arrangement are more conserved, making it easier to design primers and clone genes; (3) cpDNA is maternally inherited and thus without the genetic reassortment that interferes with the molecular phylogenetic relationships [4,5].

The chloroplast genome is a circular chromosome of 120~220 kb that consists of two inverted repeats (IRa and IRb), a large single-copy region (LSC), and small single-copy region (SSC). This conserved structure and sequence information provides a resource for primer design for other cpDNA sequencing by PCR [6]. This approach has been used for the sequencing of two bamboo cpDNA genomes [7]. As chloroplast genome of one member of the Orchidaceae family, Phalaenopsis aphrodite, has already been published [8], it is very useful to sequence complete cpDNA from another orchid, such as Oncidium using PCR.

The Chloroplast genome also has applications in plant biotechnology. Chloroplast genetic engineering offers a number of unique advantages, including high levels of transgene expression, multi-genie engineering in a single transformation event, transgene containment via maternal inheritance and a lack of gene silencing and position effects [9,10]. However, the lack of complete chloroplast genome sequences is still a major limitation to extending this technology. Additional information about the chloroplast genome would, thus, be of great value in advancing orchid biotechnology.

In this study, we designed primers based on the P. aphrodite cpDNA and used them to identify the cpDNA of Onc. Gower Ramsey, an important cut flower orchid. Such primers were also used to investigate the NADH dehydrogenase (nadh) gene deletion patterns in 15 members of the Oncidinae, and sequence amplified DNA regions to undertake phylogenetic analyses broadly across the angiosperms and at the species level.

Methods

Plant materials

Fifteen commercial Oncidiinae varieties were obtained from a grower (Yung Hsin Orchid nursery) in Taichung, Taiwan, including four Oncidium (Onc. Gower Ramsey, Gower Ramsey ‘Lemon heart’, Gower Ramsey ‘Sunkiss’, and Sweet Sugar ‘Million Coins’), five Beallara (Bllra. Eurostar, Peggy Ruth Carpenter ‘Morning Joy’, Marfitch ‘Howard Dream’, Tahoma Glacier ‘Sugar Sweet’ and Smile Eri), two Odontoglossum (Odm. Margarete Holm and Violetta von Holm), two Odontocidium (Odcdm. Golden Gate, Odcdm. Wildcat ‘Garfield’), one Degarmoara (Dgmra. Flying High) and one Zelenkocidium (Zelenkocidium Little Angel). These orchids were maintained in the greenhouse at Academia Sinica, Taipei, Taiwan, and vouchers specimens were deposited at the National Natural and Science Museum, Taichung, Taiwan. Leaves from these orchids were used in this study. Details of the parents of these species are shown in Figure 1.

DNA purification, primer design and genomic PCR

The PCR strategy for sequencing the chloroplast genome was adapted from Wu et al. [7]. For the chloroplast genomic PCR analysis, total genomic DNA from greenhouse-grown plants was isolated using a urea extraction buffer system [11]. The coding regions of the P. aphrodite chloroplast genome were used as the templates for primer design. A series of overlapping DNA fragments of 2 to 3 kb were amplified using specific primers (Additional file 1). The overlaps between adjacent PCR fragments were about 200 bp. The PCR amplification program consisted of 30 cycles of at 94°C for 30 s, at 55°C for 30 s and at 72°C for 90 s. The PCR products were sequenced. DNA sequencing was carried out with the Big-Dye Terminator Cycle Sequencing kit using an ABI Prism 3,700 DNA analyzer (Applied Biosystems, Foster City, CA). All gaps were filled by designing new primers on the basis of sequences obtained from PCR products (Additional file 1). The sequences were verified by comparison with the chloroplast genome of P. aphrodite using the VectorNTI AlignX software program (vers. 7.0; Invitrogen, Carlsbad, CA; parameters: overlap: 30; identity: 0.95; cutoff score: 40).

Broad Phylogenetic analysis

Analyses of 48 species were performed using the same 61 conserved protein-coding genes analyzed in previous studies [12-15]. This set of loci was assembled from the
aligned Nexus file for 45 species that is supplemental to the paper by Hansen et al. [14]; available from http://chloroplast.cbio.psu.edu/organism.cgi. Also included were sequences from *Lemna minor* (GenBank accession NC_010109), *Joinvillea plicata* (GeneBank accessions FJ486219 - FJ486269, L01471, U21973, and AF001864), and *Hordeum vulgare* (NC_008590) to increase sampling among monocots and break up putative long branches. Gaps introduced by the alignment were excluded from phylogenetic analyses. Two phylogenetic methods were used--maximum likelihood (ML), implemented in GARLI vers. 0.951-1 [16], and maximum parsimony (MP), implemented in PAUP* vers. 4.0b10 [17]. ML analyses were run under the general time reversible model, with all parameters estimated. A heuristic search of 100 random addition replicates was conducted for the MP analyses. Nonparametric bootstrap analyses were also performed with 100 (ML) or 1000 (MP) pseudoreplicates [18]. *Ginkgo biloba* was the specified outgroup for all analyses [14].

**Contig assembly and annotation**

VectorNTI Contig Express was used to assemble contigs (parameters: overlap: 30; identity: 0.95; and cutoff score: 40). The chloroplast genome was annotated using DOGMA (Dual Organellar GenoMe Annotator) [19]. This program uses a FASTA-formatted input file of the complete genomic sequences and identifies putative protein-coding genes by performing BLASTX searches against a custom database of published chloroplast genomes. Both tRNAs and rRNAs were identified by BLASTN searches against the same database of chloroplast genomes. For genes with low sequence identity, manual annotation was performed after identifying the position of the start and stop codons, as well as the translated amino acid sequence, using the chloroplast/bacterial genetic code.

**Analysis of variability in ndh genes of 15 Oncidium varieties**

To investigate the *ndh* genes of Oncidiae, six cpDNA regions (*trnFGAA*-*ndhF*/-ndhK-*ndhH*, *trnR<sub>ACG</sub>*-*trnNGU*, *ndhF*-rpl32, *cssA*-ndhD, *psaC*-ndhE-*ndhG*, *ndhG*-ndhI-*ndhA*-ndhH and ndhB) were obtained by a PCR approach from the 15 varieties as indicated in Methods (Accession no.: GU175359-GU175415, Additional file 2). The primer sequences, sequence size and sequence position in *Onc*. Gower Ramsey of these regions in *Onc*. Gower Ramsey are shown in Figure 2.
Phylogenetic analysis of 15 Oncidium varieties

To investigate the phylogenetic relationships between Oncidinae at the species level, eight cpDNA regions [intergene region (trnF-GUC-psbA, trnFGAA-ndhF, ycf1-trn-RACG) and coding regions (accD, matK, rbcL, rpoB, and rpoC1)] were obtained by PCR from plastid DNA of the leaves of the 15 varieties as above (Accession no.: GU915119-GU915133; GU132947-132992; GU136249-GU136275; GU175340-GU175358. Additional file 2). The primer sequences, sequence size and sequence position of these regions in Onc. Gower Ramsey are shown in Figure 2. Phylogenetics were conducted using MEGA4 (gap opening penalty: 15; gap extension penalty: 6.66; DNA weight matrix: IUB; transition weight: 0.5; negative opening penalty: 15; gap extension penalty: 6.66; DNA gene identification; 16S rRNA gene identification; 23S rRNA gene identification) were obtained by PCR from plastid DNA of the leaves of 15 varieties as above (Accession no.: GU136249-GU136275; GU175340-GU175358. Additional file 2). The primer sequences, annealing site of the forward primer in the intron of the accD gene is not included in the IR. In contrast with the chloroplast genome of Poaceae, Onc. Gower Ramsey contained introns in the rplP and rpoC1 loci and had intact copies of the accD and ycf2 genes, which are incomplete or entirely missing in Poaceae.

The broad phylogenetic analysis resulted in two trees, an ML tree with -lnL = 412281.26 (Figure 4) and an MP tree of 75,521 steps and 14,974 parsimony informative characters. The MP tree had a a consistency index (excluding uninformative characters) of 0.3649 and a retention index of 0.5997 (tree not shown). The topology of the monocot subtrees was identical for the two analyses in which Phalaenopsis was maximally supported as the sister of Oncidium and the two orchids were united with Yucca, another representative of Asparagales, with maximum support.

Analysis of variability in ndh genes of 15 Oncidium species

Six cp DNA regions (trnFGAA-ndhJ-ndhK-ndhC, trnRACG-, trnFGU-ndhF-rpl32, ccsA-ndhD, psaC-ndhE-ndhG, and ycf1-trnRACG) were obtained by PCR from plastid DNA of the leaves of 15 varieties as above (Accession no.: GU136249-GU136275; GU175340-GU175358. Additional file 2). The primer sequences, annealing site of the forward primer in Onc. Gower Ramsey and the anticipated amplicon size (bp) are presented. Different background colors indicate different experiments; gray: ndh gene identification; yellow: phylogenetic analysis.

Figure 2 Primers for Oncidium ndh gene and phylogenetic analysis. 1Primer sequences, annealing site of the forward primer in Onc. Gower Ramsey and the anticipated amplicon size (bp) are presented. 2Different background colors indicate different experiments; gray: ndh gene identification; yellow: phylogenetic analysis.
Figure 3 Gene map of Onc. Gower Ramsey chloroplast genome. The thick lines indicate the extent of the IRa and IRb, which separate the genome into SSC and LSC regions. Genes on the outside of the map are transcribed clockwise and genes on the inside of the map are transcribed counterclockwise.
Figure 4 Maximum likelihood phylogram for 61 conserved protein-coding genes. All nodes have 100% ML bootstrap support unless otherwise indicated. Horizontal branch lengths are proportional to the number of inferred substitutions/site along that branch. One node, marked “nr,” was not resolved in the ML bootstrap consensus tree. The position of Oncidium in a clade of Asparagales is indicated with an arrow.
ndhG-ndhI-ndhA-ndhH and ndhB) were obtained by PCR from total DNA of the leaves of the 15 varieties. Most of the ndh genes in the 15 Oncidiinae varieties, with the exception of ndhE in some species, had no function (Figure 5). In all 15 of the Oncidiinae varieties studied, the ndhI gene was truncated (partial sequence remained) and the ndhK gene was absent (no sequence exists). In ndhC, a frame shift occurred, creating a stop codon in the middle of the gene in all Oncidiinae, including Onc. Gower Ramsey, resulting from a 17 bp deletion (Figure 6A, Figure 5).

The ndhB of Oncidiinae does not function due to a stop codon in the first exon. To date, six orchid ndhB genes, including P. aphrodite, have been cloned and published in the NCBI database. Two of them could translate putative functional ndhB protein [Orchis rotundifolia (Accession no.: AY147484) and Coelogyne crisata (Accession no.: AY147475)]. There is also a frame shift in the second exon of Cyprpidium passeriunum (Accession no.: AY147479, AY147478.1). That of Odontoglossum crispum (AY834278) is only a partial sequence that could translate a putative ndhB protein.

The ndhf locus, which is located in the LSC-IRA junction, was absent in the Oncidiinae varieties. Notably, the nucleotide deletions in the trnRAGC-trnAGSLU-ndhf-rpl32 region were different between P. aphrodite and Oncidiinae (Figure 6B, Figure 5).

All of the 12 Oncidiinae ndhD genes cloned here were truncated. The overall pattern of truncation can be classified into two types: a truncation occurring at the 3'-end of ndhD (as in the Blira. varieties) and a truncation in the 5'-end (in the rest of the clones tested) (Figure 6C, Figure 5).

The sequences of Onc. Gower Ramsey varieties, Onc. Sweet Sugar and Odm. Margarete Holm indicate that the translation capacity of ndhE is retained in these species (Figure 5). Of the species with modified ndhE genes, Odm. Violetta. von Holm, Odcdm. Golden Gate and Odc. Wildcat contained frame shifts; Dgmsra. Flying High had a 30 bp deletion; and there were deletions of over 30 bp in the four Beallara varieties (Figure 6D, Figure 5).

Although nine varieties had no deletions in the ndhG genes, these varieties had three internal stop codons within ndhG, rendering ndhG inactive (Figure 5). There was a deletion of about 250 bp in the ndhG gene of the Beallara species (Figure 6D, Figure 5).

The region encompassing ndhl is the most complicated of the chloroplast ndh gene regions. Generally, genes from the same genus had the same pattern (for example, see Oncidium and Beallara, Figure 6E, Figure 5). In the Onc. Gower Ramsey varieties, the ndhl gene was partially deleted, and in the Beallara varieties, Zelenkocidium Little Angel and Odm. Violetta. von Holm, the ndhl gene was completely absent.

Truncated ndhA-ndhH genes still existed in most of the Oncidiinae species in this study. With the exception of frame shifts in Oncdm. Garfield ndhA and Oncdm. Golden Gate ndhH, the other ndhA through ndhH genes in the other five genera all showed deletions of various types.

![Figure 5 Summary of ndh gene patterns in Oncidiinae](image-url)

1Different background colors indicate different genera; yellow: Oncidium, white: Beallara, blue: Odontoglossum, pink: Odontocidium, purple: Colmanara, green: Degarmoara, red: Zelenkocidium. 2GR: Gower Ramsey, Sunkiss: Gower Ramsey 'Sunkiss', L. H: Lemon heart, M. C: Sunkiss 'Million Coin', E: star: Eurostar, M. J: Peggy Ruth Carpenter 'Morning Joy', H. D: Marfitch 'Howard Dream', S. S: Tahoma Glacier 'Sugar Sweet', S. E: Smile Eri, M. H: Margarete Holm, V. v.H: Violetta von Holm, G.G: Golden Gate, W.G: Wildcat 'Garfield', Dgmsra: Dgmsra. Flying High, L. A: Little Angel. 3Black star: absent genes (no sequence exists). 'White star': stop codon (There is no change in gene size but there are stop codons within coding sequences). 'Black triangle': truncated genes (only partial coding sequences are observed). 'White triangle': frame shift (reading frame shifted or nucleotides deleted). 'White circle': functional protein. - no PCR product obtained using the primers in Figure 2.
Phylogenetic analysis of 15 Oncidium species

Based on the amount of variation in the cpDNA and congruence with parent relationships, certain chloroplast regions were determined to be more useful than others. Because \textit{rbcL} is highly conserved; bootstrap scores are lower than 50% and are not useful for determining parent relationship (Additional file 3). Using the \textit{accD} gene, only the species belonging to \textit{Beallara} and \textit{Oncidium} could be separated as the pattern and relationships among other species were not correlated with the parent relationship (Additional file 4).

In the \textit{matK} region, the phylogenetic analysis of these sequences and 15 economic varieties gave results that correlated with parent relationship (Figure 7A). Therefore, we combined the most diverse cpDNA regions, \textit{trnH-psbA} [21], \textit{matK} and \textit{trnF}GAA-ndhJ [22], for phylogenetic analysis. The \textit{trnF}_{\text{GUG}}-psbA and ndhJ combination provided the most similar results to those obtained from all eight cpDNA regions (Figure 7).

Discussion

Using PCR to sequencing Oncidium Gower Ramsey and phylogenomic applications

Although there are many methods for cp genome sequencing, PCR is one method that is easy and economical [7]. However, the gene content and order in monocot cp genomes is relatively diverse so that the use of the \textit{P. aphrodite} as a template for primer design in this study was limiting. This was especially true in the \textit{ndh} gene regions where the deletion of \textit{ndh} genes in \textit{Oncidium} is very different from that in \textit{P. aphrodite}. Furthermore, when using PCR methods with total genomic DNA as the template, some of the cp sequence regions are similar to those in other organelles, thus raising the possibility of
false results. To prevent such results, we used BLAST analysis and different combinations of primers to amplify the same region.

Considerable effort is being expended to investigate phylogenomic relationships among monocots using cp genomes (see http://www.botany.wisc.edu/monatol/). Here, the phylogenetic position of Orchidaceae among Asparagales is confirmed with the robust support provided by many informative cpDNA characters. Further sampling among orchids in the future phylogenomic studies building on our results will clarify the complex relationships within the large family. Therefore, the cpDNA of *Oncidium Gower Ramsey* provides valuable information for further orchid cp genome sequencing and phylogenomics.

**ndh genes in Oncidinae cpDNA**

In higher plant chloroplasts, the NAD(P)H dehydrogenase (NDH) complex functions in PSI cyclic electron flow and chlororespiration [23]. Eleven subunits of the chloroplast *ndh* genes (*ndhA-ndhK*) are encoded in the chloroplast genome. In addition 3 cyanobacterial orthologs, nuclear-encoded subunits genes (*NdhM-NdhO*), have also been identified in chloroplasts [24]. This indicates that nucleus-encoded *ndh* genes originated in cyanobacteria and were transferred from the chloroplast genome to the nuclear genome during evolution [25]. However, in *Onc. Gower Ramsey*, out of all the 11 chloroplast-encoded *ndh* genes, only *ndhE* theoretically translates into a functional protein. This *ndh* gene truncation and absence was also observed in *P. aphrodite* [8]. Using a PCR approach to sequence the *ndh* genes of 15 varieties, we demonstrated that truncation and absence of *ndh* genes from the cp is a general phenomenon in Oncidinae.

The loss-of-function of *ndh* genes or other chloroplast-encoded genes occurs in many plants, such as parasitic plants [26-30] and achlorophyllous orchids [31,32]. Loss-of-function in *ndh* genes occurs not only in heterotrophic species, but also in autotrophic species. In *Pinus thunbergii*, all 11 *ndh* genes were putative loss-of-function alleles [33], and in another Coniferales species, *Keteleeria davidiana*, was also found to contain nonfunctional *ndh* genes [34]. In three Gnetophytes, which comprise three related families of woody gymnosperms (*Welwitschia mirabilis*, *Ephedra equisetina*, and *Gnetum parvifolium*), all 11 *ndh*

---

**Figure 7** Maximum parsimony phylogenetic trees using different cpDNA regions of 15 varieties of Oncidinae. These trees are based on the nucleotide sequences of (A) matK (B) trnH-GUG-psbA+matK (C) trnH-GUG-psbA+trnF-GAA-ndhJ (D) from all eight cpDNA regions. The numbers at the nodes indicate bootstrap support values. The scale bar indicates a branch length corresponding to 100 character-state changes.
genes are non-functional, 10 being absent and one, \textit{ndhB}, being a pseudogene [34,35]. It is interesting to note that this \textit{ndh} deletion does not occur in all gymnosperm species. The \textit{ndh} genes exist in the chloroplast genomes of \textit{Cryptomeria} and \textit{Cycas} [36,37]. It is possible that ancestral plastid \textit{ndh} genes were transferred to the nucleus, remaining functional to this day [8,33].

Loss-of-function \textit{ndh} genes also occur in other orchids [38]. \textit{Phaelenopsis aphrodite} lacks the \textit{ndhA}, \textit{ndhF}, and \textit{ndhH} genes, and only remnants of the other eight subunits sequences were found [8]. The 11 \textit{ndh} genes were either truncated or frame-shifted, suggesting that they are nonfunctional [8]. In this report, we demonstrated that \textit{ndh} gene deletion is also common in Oncidinae: the deletion pattern differs not only between \textit{Oncidium} and \textit{Phalaenopsis} (Figure 6), but even within the 15 Oncidium species analyzed (Figure 5).

From a physiological view, since parasitic plants obtain organic nutrients from the host, loss of functional \textit{ndh} genes from the chloroplast is not surprising. However, this does not explain why most \textit{ndh} genes are non-functional or deleted in autotrophic plants. The presence of \textit{ndh} homologs encoded within the nucleus was confirmed using PCR assays of total DNA of \textit{Phalaenopsis} [8]. The resulting sequences are in frame and imply that the ancestral functional \textit{ndh} copies of the plastid genome may have been transferred to the nuclear genome [8].

**Phylogenetic analysis of 15 Oncidium species**

Because it is easy to perform interspecific or intergeneric crosses with orchids, there are many artificial intergeneric hybrids. These hybrids are not distinct phenotypically and are partially named according to their parental background. However, hybrids with different parental backgrounds may be classified into the same genus. In addition, Hybrids from differently named genera may originate from the same female parent. Economic varieties of orchids are generally hybrids of other hybrids and some of the parental information has been lost. To further complicate matters, changes in the names of genera and taxonomy of the Oncidinae are frequent. In 2004, the names of more than 2200 hybrids comprising some 20% of the \textit{Oncidium} group were changed. For example, \textit{Colmanara} Wildcat was changed to \textit{Odcdm}. Wildcat and \textit{Oncidium} Little Angel was changed to \textit{Zelenkocidium} Little Angel. These changes and whether there were grounds for them could be clarified by looking carefully at the cpDNA, which could identify the female parent.

Among the eight sequences studied here, the phylogenetic analysis using \textit{matK} was most well-correlated with the parent relationship (Figure 7A). There are at least three advantages of using the \textit{matK} region for phylogenetic analysis: (1) this region is variable at the interspecies level [22]; (2) this region is easy to amplify using published primer sequences [39]; and (3) a large amount of sequence information about Oncidinae \textit{matK} is readily available in the public domain, including the number of sequences (695) and the length of the sequences (791 bp). Here, we performed a phylogenetic analysis by using 15 varieties and their 180 related sequences. Among the results we found several areas of divergence between the taxonomy of Oncidinae based on morphology and our phylogenetic analyses. For example, the female parent of \textit{Beallara} is \textit{Miltasia}, making the grandparent \textit{Brassia}. The sequences of \textit{Beallara} were highly correlated with other \textit{Brassia} species, and most closely with the female parent \textit{Brassia verrucosa} (Accession no.: EF079203, data not shown). However, the phylogenetic analysis of these sequences showed that the \textit{Odontoglossum} \textit{matK} was dispersed around the \textit{Oncidium} group (data not shown). Result such as these suggests that analysis of a single region may not contain enough information for interspecies phylogenetic analysis.

To solve this problem, available sequence information must be increased. During phylogenetic analysis, correlation is dependent on the length and properties of DNA or amino acid information. Because the information on orchid cpDNA is limited, the combination of several sequences derived by PCR using universal primers could be a successful strategy [see 22,31,32,40,41]. In this report, eight sequences from each species were combined (total length of 7042 bp) and were well-correlated with the parent relationship. However, to manage labor and supply costs, we wanted to identify the smallest region that would result in the same performance as using all eight regions. Therefore, we combined divergent cpDNA sequences such as \textit{matK} for further analyses. In addition to \textit{matK}, the \textit{trnH\textsubscript{GUG}-psbA} region is another divergent cpDNA region useful for phylogenetic analysis [21,22]. Various expansions or contractions of inverted repeats (IRs) in chloroplast genomes lead to diverse \textit{trnH\textsubscript{GUG}}, \textit{psbA} regions [42-44]. The structural changes in cpDNA provide useful phylogenetic inferences [45]. According to these data, the \textit{trnF\textsubscript{GUG}-psbA} regions are information-rich and could be used for phylogenetic analysis.

In addition, \textit{ndh} gene deletion is a unique feature that may also provide useful information for parentage analysis. The \textit{trnFGAA}, \textit{ndhJ-ndhK-ndhC} region could be amplified by PCR in all of the 15 varieties. Therefore, different combinations of these information-rich regions (\textit{matK}, \textit{trnF\textsubscript{GUG}-psbA} and \textit{trnFGAA-ndhJ}) were used for phylogenetic analysis. According to our results, two variable cpDNA regions, \textit{trnF\textsubscript{GUG}-psbA} and \textit{trnFGAA-ndhJ}, could provide sufficient information for genus-to-species level phylogenetic analysis.

However, several questions require further investigation. The first is the placement of \textit{Odm.} Violetta von Holm, whose female parent is \textit{Odm. bictoniense}. Irrespec-
tive of the cpDNA template, Odm. Violetta von Holm did not correlate with Odecim. Golden Gate or Odm. Margarete Holm, which are both derived from the same female parent. The second is the placement of Dgmra. Flying High, which has the female parent Mtsaa. Jet Setter. Theoretically, the cpDNA of Dgmra. Flying High should be closely related to Beallara species, which are derived from a Millassia female parent; however our data indicated that Dgmra. Flying High is more similar to Odontoglossum.

There are many advantages to using cpDNA for phylogenetic and parentage analysis. But this genetic information is only derived from the female parent. Therefore, in the future nuclear genes also need to be analyzed for parentage analysis [46]. In Pleione, the nrITS region was found to be more variable than the plastid regions sequenced, and nrITS gene trees were largely congruent with those inferred from the plastid regions [46]. Our data here suggest that the taxonomy of the Oncidinae may be improved by both chloroplast and nuclear genome analysis.

Conclusion
In this report, we used P. aphroide as a prototype to design primers to complete the Oncidium Gower Ramsey genome sequence. The primers and the genome sequence information obtained will be useful for further orchid cpDNA sequencing and broad phylogenetic analyses among monocots. Gene annotation showed that most of the ndh genes in Oncidinae are non-functional, with the exception of ndhE, which could theoretically produce a functional protein. In the previous reports, non-functionality of ndh genes has been found in photosynthetic orchids and gymnosperms, such as in Pinus thunbergii and Phalaenopsis. In this report, using a PCR approach, we identified the ndh genes in different Oncidium plants. The ndh genes were also non-functional in most of the plants tested, except for ndhE in four Oncidium species and Odm. Margarete Holm. These genes would be useful for parentage analysis. The non-protein coding regions trnHUG-psbA and trnFGAA-ndhJ were also determined to be cpDNA regions that would be the most useful for phylogenetic analysis. When these regions were checked in commercial varieties, most confirmed to previously known inheritance information; however, some variations need further investigation. Also, to confirm and complement the results obtained from cpDNA, genetic information may also be derived from nuclear DNA. We conclude that complete chloroplast genome information is useful for plant phylogenetic and evolutionary studies in Oncidium breeding and variety identification.

Additional material

Additional file 1 Primers for Oncidium Gower Ramsey chloroplast sequencing. Excel file containing Primers for Oncidium Gower Ramsey chloroplast sequencing.

Additional file 2 The accession numbers of the Oncidinae chloroplast sequences for ndh gene and phylogenetic analysis. Excel file containing the accession numbers of the Oncidinae chloroplast sequences for ndh gene and phylogenetic analysis.

Additional file 3 Maximum parsimony phylogenetic trees using rbcL regions of 15 varieties of Oncidinae. These trees are based on the nucleotide sequences of rbcL sequences. The numbers at the nodes indicate bootstrap support values. The scale bar indicates a branch length corresponding to 100 character-state changes.

Additional file 4 Maximum parsimony phylogenetic trees using accD regions of 15 varieties of Oncidinae. These trees are based on the nucleotide sequences of accD sequences. The numbers at the nodes indicate bootstrap support values. The scale bar indicates a branch length corresponding to 100 character-state changes.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
FHW, MTC, DCL, and CTH performed PCR and primer design. YML performed the bioinformatic analysis. HD contributed to chloroplast genome annotation, correcting errors in genome sequence, assembling the genome map and participated in manuscript preparation. MRD performed broad phylogenetic analyses and participated in manuscript preparation. CSL conceived this project, supervised PCR, primer design, bioinformatic analysis, and participated in the preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We thank Tze-in Yeh for her assistance in PCR. The authors would also like to acknowledge Dr N.D. Singh in the Daniell lab for drawing the map in Figure 3. This work was supported by the Development Program of Industrialization for Agricultural Biotechnology, Taiwan. Support for MRD was also obtained from the Plant Molecular Biology Center, Northern Illinois University, USA.

Author Details

Acknowledgements

Additional references
Webster P. Oncidium subtribe. In The Orchid Genus Book Hunters Breeze, USA, 1992.
Chase W, Palmer JD. Chloroplast DNA systematics of lilioid monocots: resources, feasibility, and an example from the Orchidaceae. Amer J Bot 1989, 76:1720-1730.
Tsai CC, Huang SC, Huang PL, Chen YS, Chou CH. Phenetic relationship and identification of subtribe Oncidinae genotypes by random amplified polymorphic DNA (RAPD) markers. Sci Hort 2002, 101:315-325.
Tien X, Li DZ. Application of DNA sequences in plant phylogenetic study. Acta Bot Yunnan 2002, 24:170-184.
Gao Q-q, Zhen Z, Jiang J, Liu Y-H, Feng Y-q, Qin L. Chloroplast DNA analysis technology and its application in Castanea. J Fruit Sci 2008, 25:396-399.
Dhingra A, Folta KM. ASAP: Amplification, sequencing, and annotation of plastomes. BMC Genomics 2005, 6:176.
Wu FH, Kan DP, Lee SB, Daniell H, Lee YW, Lin CC, Lin NS, Lin CS. Complete nucleotide sequence of Dendrocalamus latiflorus and Bambusa oldhamii chloroplast genomes. Tree Physiol 2009, 29:847-856.
8. Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Chen CH, Liu SM, Chang CC, Lin CY, Chaw SM: The chloroplast genome of Phalaenopsis aphrodite and comparative analysis of evolutionary rate with that of Grasses. *Mol Biol Evol* 2006, 23:279-291.

9. Danell H, Singh ND, Mason H, Streetfield SJ: Plant-made vaccine antigens and biopharmaceuticals. *Trends Plant Sci* 2009, 14:669-679.

10. Verma D, Daniell H: Chloroplast vector systems for biotechnology applications. *Plant Physiol* 2007, 141:129-1143.

11. Sheu JJ, Yu TS, Tong WF, Yu SM: Carbohydrate starvation stimulates differential expression of rice o-amylose genes that is modulated through complicated transcriptional and posttranscriptional processes. *J Biol Chem* 1996, 272:26998-27004.

12. Cai Z, Penafior C, Kuehl J, Leebens-Mack J, Carlson J, dePamphilis C, Boore J, Jansen R: Complete chloroplast genome sequences of Drimys, Lithiodendron, and Pipera: Implications for the phylogeny of magnoliids. *BMC Evol Biol* 2006, 6:177.

13. Goremykin VV, Hirsch-Ernst KI, Wolff S, Hellwig FH: Analysis of the Amborella trichopoda chloroplast genome sequence suggests that Amborella is not a basal angiosperm. *Mol Biol Evol* 2003, 20:1499-1505.

14. Hansen DR, Dastidar SG, Cai Z, Penafior C, Kuehl JV, Boore JL, Jansen RK: Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: Buxus (Buxaceae), Choranthus (Choranthaceae), Dioscorea (Dioscoreaceae), and Illicium (Schisandraceae). *Mol Phylogenet Evol* 2007, 45:547-563.

15. Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW: Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol Biol Evol* 2005, 22:1948-1963.

16. Zwickl DJ: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. In Ph.D. dissertation The University of Texas at Austin, 2006.

17. Swofford DL. PAUP*: Phylogenetic analysis using parsimony (*and other methods). In Version 4 Sinauer Associates, Sunderland, MA, 2003.

18. Felsenstein J: Confidence limits on phylogenies: an approach using the bootstrap. *Evolutionary Biol*, 1985, 39:783-791.

19. Wyman SK, Jansen RK, Boore JL: Automatic annotation of organelar genomes with DOGMA. *Bioinformatics* 2004, 20:3252-3255.

20. Tamura K, Dudley J, Nei M, Kumar S: MEGA 4: molecular evolutionary genetics analysis (MEGA4) software Version 4.0. *Mol Biol Evol* 2007, 24:1596-1599.

21. Kress WJ, Wurdack KJ, Zimmere EA, Wiegert LA, Jansen DH: Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* 2005, 102:8369-8374.

22. Lahaye R, Bank M van der, Bogarin B, Warner J, Pupulin F, Gigot G, Maurin TW, Boore JL, Jansen RK, dePamphilis CW: Isolation of a complex I analogue from plastid Ndh complex functioning in higher plants. *Proc Natl Acad Sci USA* 2005, 102:200-205.

23. Rumeau D, Bécuwe-Linka N, Beyly A, Louwagie M, Garin J, Peltier G: Functional gene losses occur for plastid Ndh complex functioning in higher plants. *Plant Cell* 2005, 17:219-232.

24. Muraoka R, Okuda K, Kobayashi Y, Shikani T: A eukaryotic factor required for accumulation of the chloroplast NAD(P)H dehydrogenase complex in Arabidopsis. *Plant Physiol* 2006, 142:1683-1689.

25. dePamphilis CW, Palmer JD: Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 1990, 348:337-339.

26. Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Ploock SA, Wolf PG, dePamphilis CW, Boore JL: Goffinet B: Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Mol Biol Evol* 2008, 25:393-401.

27. Haberhausen G, Zetsche K: Functional loss of ndh genes in an otherwise unaltered plastid genome of the holoparasitic flowering plant *Cuscuta reflexa*. *Plant Mol Biol* 1994, 24:217-222.

28. McNeal JR, Kuehl JV, Boore JL, dePamphilis CW: Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol* 2007, 7:57.