The complete chloroplast genome of *Dendrobium nobile*, an endangered medicinal orchid from northeast India and its comparison with related *Dendrobium* species

Ruchishree Konhar¹, Manish Debnath², Santosh Vishwakarma², Atanu Bhattacharjee², Durai Sundar³, Pramod Tandon⁴, Debasis Dash Corresp.¹, Devendra Kumar Biswal Corresp.²

¹ G.N. Ramachandran Knowledge Centre for Genome Informatics, Academy of Scientific and Innovative Research, CSIR-Institute of genomics and Integrative Biology, Delhi, India
² Bioinformatics Centre, North-Eastern Hill University, Shillong, Meghalaya, India
³ Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology, Delhi, India
⁴ Biotech Park, Lucknow, Uttar Pradesh, India

Corresponding Authors: Debasis Dash, Devendra Kumar Biswal
Email address: ddash@igib.res.in, devendra@nehu.ac.in

The medicinal orchid genus *Dendrobium* belonging to the Orchidaceae family is a huge genus comprising about 800-1500 species. To better illustrate the species status in the genus *Dendrobium*, a comparative analysis of 33 available chloroplast genomes retrieved from NCBI Refseq database was compared with that of the first complete chloroplast genome of *D. nobile* from north-east India based on next-generation sequencing methods (Illumina HiSeq 2500-PE150). Our results provide comparative chloroplast genomic information for taxonomical identification, alignment-free phylogenomic inference and other statistical features of *Dendrobium* plastomes, which can also provide valuable information on their mutational events and sequence divergence.
The Complete chloroplast genome of Dendrobium nobile, an endangered medicinal orchid from northeast India and its comparison with related Dendrobium species.

Ruchishree Konhar$^{1,2#}$, Manish Debnath$^1$, Santosh Vishwakarma$^1$, Atanu Bhattacharjee$^1$, Durai Sundar$^3$, Pramod Tandon$^4$, Debasis Dash$^{2*}$, Devendra Kumar Biswal$^{1*}$

1 Bioinformatics centre, North-Eastern Hill University, Shillong, Meghalaya, India
2 G.N. Ramachandran Knowledge Centre for Genome Informatics, Academy of Scientific and Innovative Research, CSIR-Institute of genomics and Integrative Biology, Delhi, India
3 Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology, Delhi, India
4 Biotech Park, Lucknow, Uttar Pradesh, India

#Current affiliation
*Corresponding author (s)

CORRESPONDING AUTHORS
Debasis Dash$^{2*}$
Email: ddash@igib.res.in
Devendra Kumar Biswal$^{1*}$
Email: devbioinfo@gmail.com

Abstract
The medicinal orchid genus Dendrobium belonging to the Orchidaceae family is a huge genus comprising about 800-1500 species. To better illustrate the species status in the genus Dendrobium, a comparative analysis of 33 available chloroplast genomes retrieved from NCBI Refseq database was compared with that of the first complete chloroplast genome of D. nobile from north-east India based on next-generation sequencing methods (Illumina HiSeq 2500-PE150). Our results provide comparative chloroplast genomic information for taxonomical identification, alignment-free phylogenomic inference and other statistical features of
Dendrobium plastomes, which can also provide valuable information on their mutational events and sequence divergence.

Introduction

Dendrobium is a huge genus of the tribe Dendrobieae (Orchidaceae: Epidendroideae) that was established by Olof Swartz in 1799. It includes approximately 800-1500 species and occurs in diverse habitats throughout much of Southeast Asia, including China, Japan, India, and the Philippines, Indonesia, New Guinea, Vietnam, Australia and many of the islands in the Pacific [1].

Many species and cultivars of this genus are well-known floral motifs and have featured in artwork. Dendrobium orchids are popular not only for their visual appeal in cut flower market, but also for their herbal medicinal history of about 2000 years in east and south Asian countries [2]. Many species in this genus have been extensively used as herbal medicines for several hundreds of years in treating diseases like kidney and lung ailments, gastrointestinal problems, lumbago and arthralgia. The plant extracts are also used as tonic for strengthening body’s immunity and improving sexual potency. However, many Dendrobium species in the wild face an extreme threat of extinction due to their low germination and slow growth rate, habitat decline and over exploitation arising out of anthropogenic activities [3].

Dendrobium orchids have overwhelmed researchers because of their high economic importance in global horticultural trade and in Asian traditional medicine leading to extensive plant systemic studies particularly in species identification, novel marker development, breeding and conservation. In the past two decades, promising advances have been made in areas of molecular taxonomy, plant systematics and selective breeding of Dendrobium species by intensive use of molecular markers. Recently, a variety of molecular markers like microsatellite (SSR), Random Amplified Polymorphic DNA (RAPD) and Amplified Fragment Length Polymorphism (AFLP) markers including several other DNA barcode markers from different loci of nuclear and chloroplast (cp) regions have been developed to study Dendrobium diversity. However, these species are notoriously difficult to identify [4].

The complete chloroplast (cp) genome usually contains a uniparentally inherited DNA, a feature which makes it an obvious choice for plant taxonomical analyses, phylogenomics and phylogeographic inferences at different taxonomic levels. One such classic example is the study
of phylogenetic relationships among all families in the Order Liliales, based on 75 plastid genes from 35 species in 29 genera and 100 species spanning all monocot and major eudicot lineages, where underlying results were calibrated against 17 fossil dates to redefine the monocot evolutionary timelines [5]. The significance of plastome-scale data was very well demonstrated in another study that highlighted a new functional model for understanding monocot evolution and some of their derived morphological features by way of convergent evolution from submersed aquatic ancestors (aquatic Hydatellaceae) [6]. The evolution of orchids, the largest and most diverse family of flowering plants second only to Asteraceae on Earth has long puzzled Charles Darwin and many other scientists. Recent advances in chloroplast genomics are giving researchers insights into the evolutionary history of these plants. One such study hypothesizes orchids to have arisen in Australia 112 Ma followed by migration to the Neotropics via Antarctica by 90 Ma. With the use of a combination of plastid genes, it was established that orchids and epidendroids exhibited maximally accelerated net diversification in Southeast Asia and the Neotropics respectively [7].

Studies pertaining to plastome genome sequences are useful in investigating the maternal inheritance in plants, especially those with polyploid species, owing to their high gene content and conserved genome structure [8,9 & 10]. Many species of orchids and other flowering plants exhibit rapid evolution and diversity. One of the main reasons for such diversity can be attributed to allopolyploidy or genetic redundancy, in which there are more than one gene involved in performing a particular task. In cases of useful mutation, plants evolve into new species. Hybridization and polyploidy are the decisive forces behind evolution and speciation. In the past there have been studies where a combination of AFLPs, cpDNA markers and flow cytometry was harnessed to investigate the evolutionary outcomes of hybridization between two endemic Ecuadorian species of Epidendrum (Orchidaceae) in three hybrid zones. The outcome of this study highlights the importance of hidden hybrid genotypes and their frequency which could help unravel the mysteries behind orchid evolution [11]. The advent of high-throughput sequencing technologies has enabled a rapid increase in the rate of completion of cp genomes with faster and cheaper methods to sequence organellar genomes [12, 13]. At the time of writing this manuscript, cp genomes from 33 Dendrobium species have been reported as per NCBI Organellar genome records (https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/dendrobium).
D. nobile Lindl. is one of the many highly prized medicinal plants in the genus Dendrobium. It is an endangered medicinal orchid listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) Appendix II that demands immediate attention for its protection and propagation. Here, we report the first complete cp genome of D. nobile from north-east India based on next-generation sequencing methods (Illumina HiSeq 2500-PE150) and further compare its structure, gene arrangement and microsatellite repeats with 33 existing cp genomes of Dendrobium species. Our results provide comparative chloroplast genomic information for taxonomical identification, phylogenomic inference and other statistical features of Dendrobium plastomes. These can give further insights into their mutational events and sequence divergence. The availability of complete cp genome sequences of these species in the genus Dendrobium will benefit future phylogenetic analyses and aid in germplasm utilization of these plants.

Materials and Methods

Sample collection, DNA extraction and sequencing
Fresh leaves of D. nobile were collected from plants growing in greenhouses of National Research Centre for Orchids, Sikkim, India and voucher specimen was deposited in Botanical Survey of India as well as in the Department of Botany, North Eastern Hill University, Shillong. The high molecular weight cp DNA was extracted using a modified CTAB buffer, and treated according to a standard procedure for next generation sequencing on Illumina HiSeq 2500-PE150. The quality and quantity of the genomic DNA was assessed through agarose gel electrophoresis, Nanodrop and Qubit detection method. The experiments included both paired-end and mate-pair libraries. Tagmentation was carried out with ~4μg of Qubit quantified DNA and the tagmented sample was washed using AMPURE XP beads (Beckman Coulter #A63881) and further exposed to strand displacement. The strand-displaced sample of 2-5kb and 8-13kb gel was size selected and taken for overnight circularization. The linear DNA was digested using DNA Exonuclease. Further the circularized DNA molecules were sheared using Covaris microTUBE, S220 system (Covaris, Inc., Woburn, MA, USA) for obtaining fragments in the range 300 to 1000bp. M280 Streptavidin beads (ThermoFisher Scientific, Waltham, MA) was used to cleanse the sheared DNA fragments with biotinylated junction adapters. The bead-DNA complex was subjected to End repair, A-Tailing and Adapter ligations.
Data processing

The data quality assessment for Illumina WGS raw reads was carried out using FastQC tool. Perl scripts were written for adapter clipping and low quality filtering. Cp genomes of *D. officinale*, *D. huoshanense* and *D. strongylanthum* retrieved from NCBI-Refseq database was used as reference for the assembly. BWA-MEM algorithm with default parameter settings was used for aligning the adapter clipped and low quality trimmed processed reads with the *Dendrobium* cp genomes [14]. SPAdes-3.6.0 program was used for k-mer based (k-mer used 21, 33, 55 and 77) de-novo assembly with the aligned reads and the quality of the assembled genome was gauged using Samtools and Bcftools (read alignment and genome coverage calculation) [15] (https://samtools.github.io/bcftools/bcftools.html). The cp genome of *D. nobile* was also generated through reference-assisted assembly using the high quality paired-end libraries by NOVOPlasty [16] for further validation. It is specifically designed for de novo assembly of mitochondrial and cp genomes from WGS data with the aid of a reference or seed sequence. The seed sequence can correspond to partial or complete sequence of chloroplasts of closely to distantly related species. The cpDNA refseq sequence of *D. officinale* was used as a seed sequence to perform reference-assisted assembly.

Genome annotation and codon usage

Basic Local Alignment Search Tool (BLAST; BLASTN, PHI-BLAST and BLASTX) [17], chloroplast genome analysis platform (CGAP) [18] and Dual Organellar GenoMe Annotator (DOGMA) [19] was used to annotate protein-coding and ribosomal RNA genes. The boundaries of each annotated gene with putative start, stop, and intron positions were manually determined by comparison with homologous genes from other orchid cp genomes. Further tRNA genes were predicted using tRNAscan-SE [20] and ARAGORN [21]. RNA editing sites in the protein-coding genes (PCG) of *D. nobile* were predicted using Plant RNA Editing Prediction & Analysis Computer Tool (PREPACT) (http://www.prepact.de). For this analysis, *D. nobile* cp genome was BLAST aligned against *Nicotiana tabacum*, *Oryza sativa Japonica* Group, *Phalaenopsis aphrodite* subsp. Formosana, *Physcomitrella patens* subsp. *patens* and *Zea mays* with a cutoff E-value set to 0.08. The circular genome map was drawn in OrganellarGenomeDRAW [22] followed by manual modification. The sequencing data and gene annotation were submitted to
GenBank with accession number KX377961. MEGA 7 was used to analyze and calculate GC content, codon usage, nucleotide sequence statistics and relative synonymous codon usage (RSCU) [23].

**Gene Ontology annotation and assignment of GO IDs**

Gene Ontology (GO) annotation of *D. nobile* chloroplast genes was carried out in Blast2go [24] by blast aligning the gene sequences from the GenBank annotation files to Orchidaceae sequences in non-redundant (nr) database with an e-value cutoff of $1e^{-5}$ and queried in InterProScan [25]. GO mapping and annotation of genes followed this from blast results and were subsequently merged with GO IDs from InterProScan. The merged GO annotations were validated based on True-Path-Rule by removing redundant child terms for each gene sequence. The GO annotations were slimmed down using plant-slim option.

**Simple sequence repeats analysis**

MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html), a tool for identification and location of perfect microsatellites and compound microsatellites was used to search for potential simple sequence repeats (SSRs) loci in the cp genomes of different *Dendrobium* species. The threshold point for SSRs identification was set to 10, 5, 4, 3, and 3 for mono-, di-, tri-, tetra-, and penta-nucleotides SSRs, respectively. All the repeats found were manually curated and the redundant ones were removed.

**Phylogenetic reconstruction with whole genome alignment and rearrangement analysis**

For phylogenetic reconstruction, we included *D. nobile* cp genomes from India and China along with 32 other *Dendrobium* cp genomes retrieved from GenBank. Four *Goodyera* species were taken as outgroup. The cp genome sequences were aligned with MAFFT v7.0.0 [26] and manually curated by visual inspection. PCGs as well as whole cp genomes were used for bayesian phylogenetic reconstruction using MRBAYES 3.2.6 [27]. To further validate our results we employed “K-mer Based Tree Construction” in CLC Genomics Workbench that uses single sequences or sequence lists as input and creates a distance-based phylogenetic tree. For visualization and testing the presence of genome rearrangement and inversions, gene synteny was performed using MAUVE as implemented in DNASTAR 12.3 with default settings.
Comparative analysis of intra nucleotide diversity ($Pi$) within the *Dendrobium* cp genomes was performed using MEGA 7.

**Single Nucleotide Polymorphism identification and phylogenetic analysis without genome alignment**

Phylogenetic tree was constructed based on the Single Nucleotide Polymorphisms (SNPs) identified in the whole cp genomes using kSNP3.0 with default settings except for k-mer size [28]. SNPs were identified with k-mer size set to 23, based on which, approximately 79% of the k-mers generated from median-length genome were unique.

**Results**

**Genome organization and features**

The complete cp genome of *D. nobile* was determined from the data generated out of a whole genome project initiative of the same species by Paired-end and Mate pair data from Illumina HighSeq with 150*2 and Illumina NextSeq500 with 75*2 respectively. Further the aligned Illumina reads were separated and assembled using CLC Main Workbench Version 7.7.1 into the single longest scaffold. The *D. nobile* cp genome is a typical circular double-stranded DNA with a quadripartite structure; it is 152,018 bp in size and consists of Large Single Copy (LSC) (1..84,944; 84,944 bp), Small Single Copy (SSC) (111,230..125,733; 14504 bp), and two Inverted Repeat (IR) regions of 26,285 bp: IRA (84,945..111,229) and IRB (125,734..152018).

In total 134 unique genes (79 PCGs, 8 rRNA genes, 7 pseudogenes and 38 tRNA genes) were successfully annotated, of which 12 genes {rps16, atpF, rpoC1, ycf3, rps12 (2), clpP, petB, rpl2 (2), ndhB (2)} are reported with introns (Fig. 1). We could identify a total of 20, 81 and 11 genes duplicated in the IR, LSC and SSC regions respectively in the *D. nobile* cp genome. There were a total of 49 RNA editing sites predicted in 23 genes of *D. nobile* cp genome. The whole chloroplast genome alignment included 34 *Dendrobium* species and 4 species from the genus *Goodyera* as outgroup. Each genome’s panel contained its name, sequence coordinates and a black coloured horizontal centre line with coloured block outlines appearing above and below it. Homology between the cp genomes is represented by each block with the genes, internally free from genomic rearrangement, connected by thin lines to similarly coloured blocks depicting comparative homology between the genomes (Fig. 2). The positions of LSC/IRA/SSC/IRB
borders revealed similar structures at the IR/LSC junction in the overall alignment of *Dendrobium* whole cp genomes (Fig. 3).

**Gene Ontology mapping and annotation**

We further analyzed the *D. nobile* coding cp genome sequences using the Blast2GO suite and annotated the sequences for three GO terms (biological process, molecular function, and cellular component). In case of GO term there were a total of 231 annotations in biological process (P) molecular Function and cellular compartment (C) level. In the category of biological processes a large number of these sequences are annotated for translation, photosynthesis, metabolic processes, and ribosome biogenesis. Similarly, for the GO term molecular function, the top GO categories include functions related to structural molecule activity, catalytic activity, ion and rRNA binding, transporter and transferase activity. Finally, terms including membrane, ribosome and thylakoid were annotated GO categories for cellular compartment with most of the sequences. These results are summarized along with the information on RNA editing sites in Table 1.

**Simple sequence repeat identification**

SSRs were identified in MISA perl scripts with a minimum of 10 bp repeats among all the *Dendrobium* species. Of all the SSRs, the mononucleotide A/T repeat units occupied the highest proportion. A higher proportion of di-, tri- repeats are reported rather than tetra- and penta-nucleotide repeats across *Dendrobium* cp genomes (Fig. 4).

**Phylogenetic analysis**

Phylogenetic analyses of chloroplast PCGs from *Dendrobium* species were performed with or without partitions of sequences. Both bayesian and K-mer based trees (Figs. 5 and 6) recovered a monophyly of the *Dendrobium* species, irrespective of whether or not the partitions of sequences were incorporated in the analysis supported by strong bootstrap values. The phylogenetic analyses based on complete cp genomes, suggested that five major subgroups within the genus *Dendrobium* evolved in a nested evolutionary relationship. *D. aphyllum*, *D. parishii*, *D. loddigesii* and *D. primulinum* are the most recently evolved species that nested into a single monophyletic sub group within the *Dendrobium* clade. *D. chrysotoxum* and *D. salaccense* were a
bit primitive on the evolutionary ladder in the phylogenetic tree. *Goodyera* species emerged as the outgroup that claded separately in the overall tree topology. Similar results were also obtained in the alignment free phylogenetic tree with SNPs (Fig. 6).

**Discussion**

**Potential RNA editing sites**

RNA editing is involved in plastid posttranscriptional regulation and thus provides an effective way to create transcript and protein diversity [29, 30]. In Orchidaceae, RNA editing sites were identified in 24 protein-coding transcripts in *P. aphrodite* [31]. Earlier studies indicate RNA editing sites from the same subfamily to be more conserved than those from different subfamily [32]. However, orchids and other angiosperms have relatively less common editing sites. For example, orchids and *Cocos nucifera* share 10 potential RNA editing sites; comparisons among *Nicotiana tabacum*, *Arabidopsis thaliana* and orchid RNA editing sites have shown low conservation of editing sites (one common editing site in *rpoB*) [32]. Our studies congruently predicted all 49 RNA editing sites (Table 1) in 23 genes of *D. nobile* from at least 75% of the reference organisms (*Nicotiana tabacum*, *Oryza sativa Japonica* Group, *Phalaenopsis aphrodite* subsp. *Formosana*, *Physcomitrella patens* subsp. Patens and *Zea mays*) and resulted in amino acid substitutions. All the RNA-editing sites were non-silent and edited C to U. Of the 49 RNA editing sites 89.8% (44) editing sites appeared in the second position of triplet codon, 10.2% (5) editing sites appeared in the first position of triplet codon whereas no editing sites appeared in the third base of triplet codon. The genes *ndhD*, *rpoB*, *rpoC1* had 8, 6 and 4 RNA editing sites respectively. All the 49 RNA editing sites led to changes in the amino acid. The most frequent amino acid conversion was hydrophilic to hydrophobic (S to L, 22 occurrences and S to F, 8 occurrences), followed by hydrophobic to hydrophobic conversions (P to L, 12 occurrences). Seven conversions were found to be hydrophilic to hydrophilic (H to Y, 5 occurrences and T to M, 2 occurrences).

**Comparison with other chloroplast genomes within the genus Dendrobium**

We compared thirty-four chloroplast genomes representing different species within the genus *Dendrobium* (Table 2). The length of the *Dendrobium* species cp genomes ranged from 148,778 to 153,953 bp, with *D. chrysotoxum* being the largest cp genome and *D. moniliforme* the
The cp genomes have acquired the familial angiosperm plastome organization comprising of a LSC, an SSC and a pair of IR regions each. *Dendrobium* cp genomes are also AT-rich (62.26–62.39%) quite similar to other orchid cp genomes [33]. Differences in the cp genome size of these species are primarily due to the variations in the length of LSC, SSC and IR regions. Synteny comparison revealed a lack of genome rearrangement and inversions, thereby, substantiating for the highly conserved nature in the genomic structure, including gene number and gene order in these cp genomes. However, structural variation was predominant in the LSC/IR/SSC boundaries (Fig. 2), which could be harnessed for predicting potential biomarkers for species identification.

IR regions are generally considered to be highly conserved regions in the chloroplast genome. IR expansion or contraction is determined by the variability of genes flanking IR/SC junctions [27]. In the evolutionary ladder, SSC and IR border regions experience expansion and contraction that overall contribute to the variation in chloroplast genome length [34, 35]. At the IR/LSC boundaries, most IRs of non-orchid monocots exhibit trnH-rps19 gene clusters, excluding Ψrpl22 genes, leading to more-progressive expansion of IRs compared to non-monocot angiosperms [36, 37]. Contrarily, the orchid chloroplast genomes have distinct characteristics at the IR/SSC junction and are classified into four types based on the organization of genes flanking the IRB/SSC junction (J\(_{SB}\)). In type I structure, J\(_{SB}\) is located upstream of the ndhF-rpl32 cluster and is primarily seen in *Cypripedium* and *Dendrobium* species. Type II junction is found in *Cymbidium* species in which J\(_{SB}\) is located within Ψycf1 and ndhF genes. Type III is reported in *Oncidium*, *Erycina*, and *Phalaenopsis equestris*, in which J\(_{SB}\) is located inside the Ψycf1-rpl32 cluster, with the loss of ndhF gene. The type IV structure is characterized by the incorporation of the entire ycf1 into the SSC, with J\(_{SB}\) inside trnN-rpl32 [28]. In the present study, the positions of LSC/IRA/SSC/IRB borders were examined in the overall alignment of *Dendrobium* whole cp genomes and all of them were found to have similar structures at the IR/LSC junction akin to type I structure (Fig. 3). Previous studies emphasize that IR expansion or contraction may not correlate with the taxonomic relationships [29]. More molecular data is required for enhancing our present understanding of the genes flanking IR/SSC junctions and their underlying variations.
A comparative nucleotide sequence statistics (counts of annotations, AT/GC counts, nucleotide frequency in codon positions etc.) for all the *Dendrobium* species including representatives from outgroup are outlined in Tables 3, 4 and 5. The relative synonymous codon usage is given in parentheses following the codon frequency (averages over all taxa) involved (Table 6). Maximum Likelihood analysis of natural selection codon-by-codon was carried out. For each codon, estimates of the numbers of inferred synonymous (s) and nonsynonymous (n) substitutions are presented along with the number of sites that are estimated to be synonymous (S) and nonsynonymous (N) (Table S1). These estimates were calculated using the joint Maximum Likelihood reconstructions of ancestral states under a Muse-Gaut model [38] of codon substitution and Felsenstein 1981 model [39] of nucleotide substitution. For estimating ML values, a tree topology was automatically computed. The test statistic dN - dS was used for detecting codons that have undergone positive selection, where dS is the number of synonymous substitutions per site (s/S) and dN is the number of nonsynonymous substitutions per site (n/N). A positive value for the test statistic indicates an overabundance of nonsynonymous substitutions. In this case, the probability of rejecting the null hypothesis of neutral evolution (P-value) was calculated [40, 41]. Values of P less than 0.05 are considered significant at a 5% level and are highlighted [Table S2]. Normalized dN - dS for the test statistic is obtained using the total number of substitutions in the tree (measured in expected substitutions per site). The analysis involved 38 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding and all positions containing gaps and missing data were eliminated. There were a total of 108,594 positions in the final dataset.

Gene ontology analysis

The GO annotation revealed majority of the chloroplast genes are involved in the process of translation, photosynthesis, ion transport and transcription (Table 1). The molecular functions of the genes are majorly binding – RNA, metal ion, DNA, ion and electron transport, RNA polymerase activity and various other enzymatic activities. Enzyme classification showed seven genes to be translocases, four as transferases, two as oxidoreductases, and one each as hydrolase, lyase and ligase. A majority of the genes encode proteins localizing in chloroplast thylakoid.
membrane, ribosome and few are transported to the mitochondria. The ndhB gene is involved in photosynthesis, while rpoB and rpoC1 are involved in biosynthetic process.

Characterization of simple sequence repeats

Previous studies have documented prevalence of mononucleotide and dinucleotide SSRs in at least 15 *Dendrobium* species from 92 syntenic intergenic and intronic loci. Of all these loci, 10 (mutational hotspots: *psbB-psbT, rpl16-rps3, trnR-atpA*, and *trnL* intron); (*ndhF-rpl32, rpl32-trnL, rps16-trnQ, and trnE-trnT*) are reported to be the fastest evolving and are termed as top ten hotspots [29]. The SSRs lying in these regions could be further investigated for identifying potential markers that can aid in barcoding analysis.

Phylogenetic analyses

In the present study, we employed two different approaches for phylogeny reconstruction. First we aligned the whole cp genomes and exported the alignment matrices for creating a Bayesian tree (Fig. 5). Two independent MCMC chains were run with first 25% of the cycles removed as burn-in, coalescence of substitution rate and rate model parameters were also examined and average standard deviation of split frequencies was carried out and generations added until the standard deviation value was lowered to 0.01. Secondly we performed a phylogenetic tree construction using an alignment free approach. In this case we identified the SNPs from the cp genomes and utilised them in constructing the phylogenetic tree (Fig. 6). A total of 13,839 SNPs were identified in the 38 genomes analyzed, of which 2,203 were homoplastic SNPs i.e. SNPs that do not correspond to any node in the parsimony tree. The fraction of k-mers present in all genomes is 0.482. The numbers at the nodes in the phylogenetic tree indicate the number of SNPs that are present in all of the descendants of that node and absent in others (Fig. 6). The numbers at the tips indicate the number of SNPs unique to each particular species.

The two different methods that employed both alignment and alignment-free approach resulted in highly reliable identical phylogenetic trees within each data set. Different analyses based on the two datasets generated largely congruent topologies (Figs. 5 and 6) with *Dendrobium* species forming one clade and *Goodyera* species forming another clade as an outgroup.

Conclusions
This study provides the first comparative account on the complete chloroplast genome of *D. nobile* from north-east India with 33 other species from the genus *Dendrobium* that revealed higher sequence variation in SSC and LSC regions compared with IR regions in both coding and non-coding regions. The gene order, gene content and genomic structure were highly conserved with those of other sequenced *Dendrobium* species. However, IR contraction is observed within the genus and several SNPs identified from these cp genomes were quite instrumental in generating alignment-free robust phylogenetic trees that congrued with trees generated from aligned matrices of whole cp genomes. This gives an indication that the SNPs, insertions and deletions, LSC and SSC regions in the cp genomes of this medicinal orchid genus can be utilized for barcoding and biodiversity studies. Further this would augment more and more plastome sequencing of *Dendrobium* species that are not reported in this study.

References
1. Wood, H. P. 2006. *The Dendrobiums*. Timber Press
2. Bulpitt CJ, Li Y, Bulpitt PF, Wang J. 2007. The use of orchids in Chinese medicine. *Journal of the Royal Society of Medicine* 100(12):558-63
3. Kong JM, Goh NK, Chia LS, Chia TF. 2003. Recent advances in traditional plant drugs and orchids. *Acta Pharmacologica Sinica* 24(1):7-21
4. Teixeira da Silva JA, Jin X, Dobránszki J, Lu J, Wang H, Zotz G, Cardoso JC, Zeng S. 2016. Advances in Dendrobium molecular research: Applications in genetic variation, identification and breeding. *Molecular Phylogenetics and Evolution* 95:196-216
5. Givnish TJ, Zuluaga A, Marques I, Lam VKY, Gomez MS, Iles WJD, Ames M, Spalink D, Modeller JR, Jackson R, Briggs BG, Lyon SP, Stevenson DW, Zomlefer W, Graham SW. 2016. Phylogenomics and historical biogeography of the monocot order *Liliales*: Out of Australia and through Antarctica. *Cladistics* 32: 581–605
6. Givnish TJ, Zuluaga A, Spalink D, Soto Gomez M, Lam VKY, Saarela JM, Sass C, Iles WJD, de Sousa DJL, Leebens-Mack J, Chris Pires J, Zomlefer WB, Gandolfo MA, Davis JI, Stevenson DW, dePamphilis C, Specht CD, Graham SW, Barrett CF, Ané C. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am J Bot.* 105(11):1888-1910
7. Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A. 2016. Orchid historical biogeography, diversification, Antarctica and the paradox of orchid dispersal. *J Biogeogr.* 43:1905–16

8. Birky CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America* 92(25):11331–11338

9. Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):7051–7057

10. Song BH, Wang XQ, Wang XR, Sun LJ, Hong DY, Peng PH. 2002. Maternal lineages of *Pinus densata*, a diploid hybrid. *Molecular Ecology* 11:1057–1063

11. Marques I, Draper D, Riofrío L, Naranjo C. 2014. Multiple hybridization events, polyploidy and low postmating isolation entangle the evolution of neotropical species of Epidendrum (Orchidaceae). *BMC Evol Biol.* 14:20

12. Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, Tomkins J, Rognli OA, Daniell H, Clarke JL. 2007. Complete chloroplast genome sequences of Hordeum vulgare, Sorghum bicolor and Agrostis stolonifera, and comparative analyses with other grass genomes. *Theoretical and Applied Genetics* 115(4):571-90.

13. Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. 2008. Multiplex sequencing of plant chloroplast genomes using Solexá sequencing-by-synthesis technology. *Nucleic Acids Research* 36(19):e122.

14. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009 Jul 15;25(14):1754-60.

15. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5):455-77.

16. Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):e18.
17. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**(17):3389-402.

18. Cheng J, Zeng X, Ren G, Liu Z. 2013. CGAP: a new comprehensive platform for the comparative analysis of chloroplast genomes. *BMC Bioinformatics* **14**:95.

19. Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**(17):3252-5.

20. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**(5):955-64.

21. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* **32**(1):11-6.

22. Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **41**(Web Server issue): W575-81.

23. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**(7):1870-4.

24. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. **21**(18): 3674-6

25. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **30**(9):1236-40.

26. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**(4):772-80.

27. Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8):754-5.

28. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* **1**(31):2877-8
29. Chen C, Bundschuh R. 2012. Systematic investigation of insertional and deletional RNA-DNA differences in the human transcriptome. *BMC Genomics* 13: 616
30. Knoop V. 2011. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci* 68: 567–586
31. Zeng WH, Liao SC, Chang CC. 2007. Identification of RNA editing sites in chloroplast transcripts of *Phalaenopsis aphrodite* and comparative analysis with those of other seed plants. *Plant Cell Physiol* 48: 362–368
32. Luo J, Hou BW, Niu ZT, Liu W, Xue QY, Ding XY. 2014. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLoS One* 9(6):e99016
33. Zhitao N, Shuying Z, Jiajia P, Ludan L, Jing S, Xiaoyu D. 2017. Comparative analysis of Dendrobium plastomes and utility of plastomic mutational hotspots. *Sci Rep.* 7(1):2073
34. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology* 8:36
35. Li R, Ma PF, Wen J, Yi TS. 2013. Complete sequencing of five araliaceae chloroplast genomes and the phylogenetic implications. *PLoS One* 8(10):e78568
36. Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssallem IS, Yu J. 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One* 5: e12762
37. Goulding SE, Olmstead RG, Morden CW, Wolfe KH. 1996. Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet* 252: 195–206
38. SV Muse, BS Gaut. (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11:715-724.
39. Felsenstein J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Molecular Evolution* 17:368-376.
40. Kosakovsky Pond, S. L., and Frost, S. D. W. (2005) Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection . *Mol Biol Evol* 22:1208-1222.
488 41. Suzuki Y, Gojobori T. (1999) A method for detecting positive selection at single amino acid sites. Mol Biol Evol 16:1315-28.
Figure Legends

**Figure 1.** Gene map of *Dendrobium nobile* chloroplast genome from Northeast India. Genes shown inside the circle are transcribed clockwise, and those outside are transcribed anticlockwise. Color coding indicates genes of different functional groups. A pair of inverted repeats (IRA and IRB) separate the genome into LSC and SSC regions.

**Figure 2.** Whole chloroplast genome alignment of 38 orchid species. The whole chloroplast genome alignment includes 34 *Dendrobium* species and 4 species from the genus *Goodyera* as outgroup. Each genome’s panel contains its name, sequence coordinates and a black coloured horizontal centre line with coloured block outlines appearing above and below it. Each block represents homology with the genes, internally free from genomic rearrangement, connected by lines to similarly coloured blocks depicting comparative homology across genomes.

**Figure 3.** Comparison of the borders of LSC, SSC and IR regions across *Dendrobium* chloroplast genomes.

**Figure 4.** SSR distribution among different *Dendrobium* plastomes. The SSR were determined in MISA per scripts based on the comparison between plastomes of each tested *Dendrobium* species and *D. nobile*. Histograms with different color codes indicate the numbers of SSRs. The minimum number (thresholds) of SSRs was set as 10, 5, 4, 3, and 3 for mono-, di-, tri-, tetra-, and penta-nucleotides SSRs, respectively.

**Figure 5.** Phylogenetic tree based on bayesian inference from the whole genome alignment matrix of *Dendrobium* chloroplast genomes. The tree yielded monophyletic groupings of the genus *Dendrobium* and *Goodyera* species emerged as outgroup with a separate clade. Posterior probability/bootstrap values are indicated on the internal nodes, which are highly supportive of the overall tree topology.

**Figure 6.** Alignment free phylogenetic tree reconstruction based on SNP identification. The optimum kmer size for the dataset is determined that calculates FCK, a measure of diversity of
sequences in the dataset (Kchooser) and a consensus of the equally most parsimonious trees are reported. The numbers at the nodes indicate the number of SNPs that are present in all of the descendants of that node and absent in others. The numbers within parentheses at the tips indicate the number of SNPs unique to each particular species.
Figure 1

Gene map of *Dendrobium nobile* chloroplast genome from Northeast India

Genes shown inside the circle are transcribed clockwise, and those outside are transcribed anticlockwise. Color coding indicates genes of different functional groups. A pair of inverted repeats (IRA and IRB) separate the genome into LSC and SSC regions.
Figure 2

Whole chloroplast genome alignment of 38 orchid species.

The whole chloroplast genome alignment includes 34 *Dendrobium* species and 4 species from the genus *Goodyera* as outgroup. Each genome’s panel contains its name, sequence coordinates and a black coloured horizontal centre line with coloured block outlines appearing above and below it. Each block represents homology with the genes, internally free from genomic rearrangement, connected by lines to similarly coloured blocks depicting comparative homology across genomes.
Figure 3

Comparison of the borders of LSC, SSC and IR regions across *Dendrobium* chloroplast genomes
Figure 4

SSR distribution among different *Dendrobium* plastomes.

The SSR were determined in MISA per scripts based on the comparison between plastomes of each tested *Dendrobium* species and *D. nobile*. Histograms with different color codes indicate the numbers of SSRs. The minimum number (thresholds) of SSRs was set as 10, 5, 4, 3, and 3 for mono-, di-, tri-, tetra-, and penta-nucleotides SSRs, respectively.
Figure 5

Phylogenetic tree based on bayesian inference from the whole genome alignment matrix of *Dendrobium* chloroplast genomes

The tree yielded monophyletic groupings of the genus *Dendrobium* and *Goodyera* species emerged as outgroup with a separate clade. Posterior probability/bootstrap values are indicated on the internal nodes, which are highly supportive of the overall tree topology.
Alignment free phylogenetic tree reconstruction based on SNP identification

The optimum kmer size for the dataset is determined that calculates FCK, a measure of diversity of sequences in the dataset (Kchooser) and a consensus of the equally most parsimonious trees are reported. The numbers at the nodes indicate the number of SNPs that are present in all of the descendants of that node and absent in others. The numbers within parentheses at the tips indicate the number of SNPs unique to each particular species.
RNA editing sites predicted in *Dendrobium nobile* chloroplast genome

*D. nobile* cp genome was BLAST aligned against reference datasets of *Nicotiana tabacum*, *Oryza sativa Japonica* Group, *Phalaenopsis aphrodite subsp. Formosana*, *Physcomitrella patens subsp. Patens* and *Zea mays*. Threshold for congruent prediction of RNA editing sites from the reference taxa was set to 3 (Count) and 75% (Percentage of prevalence). Count is in the form of (number of reference taxa against which editing site found) / (number of taxa with the homologous site).
Table 1. RNA editing sites predicted in *Dendrobium nobile* chloroplast genome along with its GO annotations.

| Gene | GO IDs | GO slim Annotation | Nucleotide Position | Amino Acid Position | Triplet position within codon | Base Conversion | Codon change | Amino Acid Conversion | Count | Percentage of Prevalence |
|------|--------|-------------------|---------------------|---------------------|-----------------------------|-----------------|--------------|----------------------|-------|------------------------|
| matK | F: GO:0005198; P: GO:0006412; C: GO:0005840; GO:0009507 | F: structural molecule activity; P: translation; C: ribosome; chloroplast | 1258 | 420 | 1 | C→U | CAC→UA | C→U | H→Y | 4/5 | 80 |
|        |        |                   | 913 | 305 | 1 | C→U | CAU→UA | U     | H→Y | 4/5 | 80 |
| rps16 | F: GO:0000166; GO:0005215; P: GO:0006139; GO:0006810; GO:0009058 | F: nucleotide binding; transporter activity; P: nucleobase-containing compound metabolic process; transport; biosynthetic process; C: chloroplast; thylakoid; membrane | 143 | 48 | 2 | C→U | UCA→UA | A     | S→L | 4/4 | 100 |
| atpA  | F: GO:0000166; GO:0005215 | F: nucleotide binding; transporter activity | 773 | 258 | 2 | C→U | UCA→UA | A     | S→L | 5/5 | 100 |
|   | GO:0006139 | GO:0006810 | GO:0009058 | P: nucleobase-containing compound metabolic process; transport; biosynthetic process |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|
| C: GO:0009507; GO:0009579; GO:0016020 |   |   |   | C: chloroplast; thylakoid; membrane |   |   |   |   |   |   |
| F: GO:0005215 | P: nucleobase-containing compound metabolic process; transport; biosynthetic process | 92 | 31 | 2 | C→U | CCA→CU | P→L | 5/5 | 100 |   |
|   |   |   |   |   |   |   |   |   |   |   |
| F: GO:0005215 | P: nucleobase-containing compound metabolic process; transport; biosynthetic process | 629 | 210 | 2 | C→U | UCA→UU | S→L | 5/5 | 100 |   |
|   |   |   |   |   |   |   |   |   |   |   |
| F: GO:0005215 | P: nucleobase-containing compound metabolic process; transport; biosynthetic process | 428 | 143 | 2 | C→U | CCU→CU | P→L | 5/5 | 100 |   |
| Gene | GO IDs | GO Terms | C→U  | UGU→UU  | S→L  | % Correct |
|------|--------|----------|------|---------|------|-----------|
| rpoC1 | GO:0005886; GO:0009507; GO:0009579 | C: plasma membrane; chloroplast; thylakoid | 617  | 206     | 2    | C→U      |
|      | GO:0003677; GO:0016740 | F: DNA binding; transferase activity | 488  | 163     | 2    | C→U      |
|      | GO:0006139; GO:0009058 | P: nucleobase-containing compound metabolic process; biosynthetic process | 182  | 61      | 2    | C→U      |
|      | GO:0009507 | C: chloroplast | 41   | 14      | 2    | C→U      |
| rpoB  | GO:0005886; GO:0009507; GO:0009579 | C: plasma membrane; chloroplast; thylakoid | 2426 | 809     | 2    | C→U      |
|      | GO:0003677; GO:0016740 | F: DNA binding; transferase activity | 623  | 208     | 2    | C→U      |
|      | GO:0006139; GO:0009058 | P: nucleobase-containing compound metabolic process; biosynthetic process | 566  | 189     | 2    | C→U      |
|      | GO:0009507 | C: chloroplast | 551  | 184     | 2    | C→U      |
| Gene | GO:0003723; GO:0005198 | GO:0005515 | GO:0006139; GO:0006810; GO:0009058 | 149 | 50 | 2 | C→U | CCA→CU A | P→L | 5/5 | 100 |
|------|------------------------|------------------------|----------------------------------|-----|-----|----|-----|--------|------|-----|-----|
| rps14 | F: RNA binding; structural molecule activity | F: protein binding | nucleotide binding; transporter activity | 191 | 64 | 2 | C→U | CCA→CU A | P→L | 5/5 | 100 |
|      | P: generation of precursor metabolites and energy; translation; photosynthesis | P: photosynthesis | P: nucleobase-containing compound metabolic process; transport; biosynthetic process | 185 | 62 | 2 | C→U | ACG→AU G | T→M | 5/5 | 100 |
|      | C: chloroplast; thylakoid; membrane; ribosome | C: chloroplast; thylakoid; membrane | C: GO:0009507; GO:0005840 | 44 | 15 | 2 | C→U | UCU→UU U | S→F | 5/5 | 100 |
| ycf3  | F: GO:0003723; GO:0005198 | F: GO:0005515 | GO:0000166; GO:0005215 | 1184 | 395 | 2 | C→U | UCA→UU A | S→L | 5/5 | 100 |
|      | GO:0006091; GO:0006412; GO:0015979 | GO:0009507; GO:0009579; GO:0016020; GO:0015979 | GO:0000166; GO:0005215 | | | | | | | |
| atpB  | | | | | | | | | | |
|      | | | | | | | | | | |
| Gene | C: GO:0009507; GO:0009579; GO:0016020 | C: chloroplast; thylakoid; membrane | F: GO:0000166; GO:0016740 | F: nucleotide binding; transporter activity | 1184 | 395 | 2 | C→U | UCA→UU A | S→L | 4/4 | 100 |
|------|--------------------------------------|--------------------------------------|-----------------------------|---------------------------------------------|------|-----|---|-----|--------|-----|-----|-----|
| accD | P: GO:0006139; GO:0006629; GO:0009058 | P: nucleobase-containing compound metabolic process; lipid metabolic process; biosynthetic process | 1412 | 471 | 2 | C→U | CCA→CU A | P→L | 3/3 | 100 |
|      | C: GO:0009507 | C: chloroplast | 1430 | 477 | 2 | C→U | CCU→CU U | P→L | 3/3 | 100 |
| psaI | P: GO:0015979; | P: photosynthesis | | | | | | | | |
|      | C: GO:0009507; GO:0009579; GO:0016020 | C: chloroplast; thylakoid; membrane | 80 | 27 | 2 | C→U | UCU→UU U | S→F | 5/5 | 100 |
| psbF | F: GO:0003824; GO:0005488 | F: catalytic activity; binding | | | | | | | | |
|      | P: GO:0006091; GO:0015979; | P: generation of precursor metabolites and energy; photosynthesis | 77 | 26 | 2 | C→U | UCU→UU U | S→F | 5/5 | 100 |
| Protein | GO IDs | GO Terms | Gene Symbols | Codon | matured | Match | Confidence |
|---------|--------|----------|-------------|-------|---------|-------|------------|
| petL    | GO:0003824; GO:0009579 | C: catalytic activity; F: catalytic activity | petL | 5 2 2 | C→U | CCU→CU | P→L | 5/5 | 100 |
| rpl20   | GO:0003723; GO:0005198 | F: RNA binding; F: structural molecule activity | rpl20 | 308 103 2 | C→U | UCA→UU | S→L | 4/5 | 80 |
| clpP    | GO:0005488 | F: hydrolase activity | clpP | 559 187 1 | C→U | CAU→UA | H→Y | 5/5 | 100 |
| petB    | GO:0003824; GO:0005488 | F: catalytic activity; binding | petB | 611 204 2 | C→U | UCA→UU | S→L | 5/5 | 100 |
| Gene | GO Terms and Functions                                                                 | C→U Sites | UCA→UU A Sites | S→L Sites | Z Values | p Values |
|------|----------------------------------------------------------------------------------------|------------|----------------|-----------|----------|----------|
| **rpoA** | C: chloroplast; thylakoid; membrane; F: DNA binding; protein binding; transferase activity | 830 | 277 | 2 | C→U | UCA→UU A | S→L | 4/4 | 100 |
| rpl2 | P: nucleobase-containing compound metabolic process; biosynthetic process | 368 | 123 | 2 | C→U | UCA→UU A | S→L | 4/4 | 100 |
| ndhD | C: chloroplast; C: ribosome; chloroplast | 200 | 67 | 2 | C→U | UCU→UU U | S→F | 3/4 | 75 |
| ndhD | F: catalytic activity; binding | 878 | 293 | 2 | C→U | UCA→UU A | S→L | 4/4 | 100 |
| ndhD | P: generation of precursor metabolites and energy | 674 | 225 | 2 | C→U | UCG→UU G | S→L | 4/4 | 100 |
| Gene | GO Terms (GO:xxxx) | Cellular Component | GO Terms (GO:yyyy) | Protein Terms (GO:zzzz) | Synonymous Substitution | C→U | S→L | Homology | Identity |
|------|-------------------|--------------------|--------------------|-------------------------|------------------------|------|------|----------|----------|
| ndhA | C: GO:0009507; GO:0009579; GO:0016020 | C: chloroplast; thylakoid; membrane | 383 | 128 | 2 | C→U | UCA→UU | 4/4 | 100 |
|      | F: GO:0003824; GO:0005488 | F: catalytic activity; binding | | | | | | |
|      | P: GO:0006091; GO:0015979 | P: generation of precursor metabolites and energy; photosynthesis | 473 | 158 | 2 | C→U | UCA→UU | 4/4 | 100 |
|      | C: GO:0005886; GO:0009507; GO:0009579 | C: plasma membrane; chloroplast; thylakoid | | | | | | |
| ndhB | F: GO:0003824; GO:0005488 | F: catalytic activity; binding | 149 | 50 | 2 | C→U | UCA→UU | 4/4 | 100 |
|      | 467 | 156 | 2 | C→U | CCA→CU | 4/4 | 100 |
|      | 586 | 196 | 1 | C→U | CAU→UA | 4/4 | 100 |
|      | 704 | 235 | 2 | C→U | UCC→UU | 4/4 | 100 |
|      | 737 | 246 | 2 | C→U | CCA→CU | 4/4 | 100 |
|      | 830 | 277 | 2 | C→U | UCA→UU | 4/5 | 80 |
|      | 836 | 279 | 2 | C→U | UCA→UU | 4/5 | 80 |
|      | 1481 | 494 | 2 | C→U | CCA→CU | 4/4 | 100 |
| rpl23 | F: GO:0003723; GO:0005198 | F: RNA binding; structural molecule activity | 71 | 24 | 2 | C→U | UCU→UU | 4/5 | 80 |
D. nobile cp genome was BLAST aligned against reference datasets of Nicotiana tabacum, Oryza sativa Japonica Group, Phalaenopsis aphrodite subsp. Formosana, Physcomitrella patens subsp. Patens and Zea mays. Threshold for congruent prediction of RNA editing sites from the reference taxa was set to ≥ 3 (Count) and 75% (Percentage of prevalence). Count is in the form of (number of reference taxa against which editing site found) / (number of taxa with the homologous site). Further the genes were exported to OMIX box, blast aligned and subsequently mapped and annotated with Gene ontology (GO) slim terms. Their corresponding GO ids and annotations are shown in the table.

| P: GO:0006412 | P: translation |
|----------------|----------------|
| C: GO:0005840; GO:0009507 | C: ribosome; chloroplast |
Table 2 (on next page)

Summary of characteristics in chloroplast genome sequences of thirty-four *Dendrobium* species and four *Goodyera* species (taken as outgroup)
Table 2. Summary of characteristics in chloroplast genome sequences of thirty-four *Dendrobium* species and four *Goodyera* species (taken as outgroup).

| Organism                  | Accession Number | Length   | Weight (single-stranded) Mda | Weight (double-stranded) Mda |
|---------------------------|------------------|----------|-----------------------------|-----------------------------|
| *Dendrobium nobile*       | KX377961         | 152018   | 46.932                      | 93.912                      |
| *Dendrobium officinale*   | NC_024019        | 152221   | 46.995                      | 94.038                      |
| *Dendrobium strongylanthum* | NC_027691   | 153059   | 47.256                      | 94.556                      |
| *Dendrobium huoshanense*  | NC_028430        | 153188   | 47.294                      | 94.635                      |
| *Dendrobium chrysotoxum*  | NC_028549        | 153953   | 47.528                      | 95.108                      |
| *Dendrobium nobile* (China) | NC_029456     | 153660   | 47.453                      | 94.927                      |
| *Dendrobium pendulum*     | NC_029705        | 153038   | 47.246                      | 94.542                      |
| *Dendrobium moniliforme*  | NC_035154        | 148778   | 45.931                      | 91.911                      |
| *Dendrobium primulinum*   | NC_035321        | 150767   | 46.545                      | 93.14                       |
| *Dendrobium aphyllum*     | NC_035322        | 151524   | 46.779                      | 93.607                      |
| *Dendrobium brymerianum*  | NC_035323        | 151830   | 46.873                      | 93.796                      |
| *Dendrobium denneanum*    | NC_035324        | 151565   | 46.793                      | 93.633                      |
| *Dendrobium devonianum*   | NC_035325        | 151945   | 46.909                      | 93.867                      |
| *Dendrobium falconeri*    | NC_035326        | 151890   | 46.891                      | 93.833                      |
| *Dendrobium gratiosissimum* | NC_035327     | 151829   | 46.873                      | 93.796                      |
| *Dendrobium hercoglossum* | NC_035328        | 151939   | 46.908                      | 93.864                      |
| *Dendrobium wardianum*    | NC_035329        | 151788   | 46.861                      | 93.77                       |
| *Dendrobium wilsonii*     | NC_035330        | 152080   | 46.951                      | 93.951                      |
| *Dendrobium crepidatum*   | NC_035331        | 151717   | 46.837                      | 93.726                      |
| *Dendrobium salaccense*   | NC_035332        | 151104   | 46.648                      | 93.347                      |
| *Dendrobium spatella*     | NC_035333        | 151829   | 46.872                      | 93.796                      |
| *Dendrobium parciiflorum* | NC_035334        | 150073   | 46.331                      | 92.711                      |
| *Dendrobium nobile*       | NC_035335        | 151850   | 46.88                       | 93.809                      |
| *Dendrobium chrysanthum*  | NC_035336        | 151790   | 46.861                      | 93.772                      |
| *Dendrobium jenkinsii*    | NC_035337        | 151717   | 46.839                      | 93.726                      |
| *Dendrobium lohobense*    | NC_035338        | 151812   | 46.868                      | 93.785                      |
| *Dendrobium parishii*     | NC_035339        | 151689   | 46.83                       | 93.709                      |
| *Dendrobium ellipsophyllum* | NC_035340    | 152026   | 46.935                      | 93.917                      |
| *Dendrobium xichouense*   | NC_035341        | 152052   | 46.942                      | 93.933                      |
| *Dendrobium fimbriatum*   | NC_035342        | 151673   | 46.825                      | 93.699                      |
| *Dendrobium exile*        | NC_035343        | 151294   | 46.707                      | 93.465                      |
| *Dendrobium fanjingshanense* | NC_035344  | 152108   | 46.96                       | 93.968                      |
| *Dendrobium candidum*     | NC_035745        | 152094   | 46.955                      | 93.959                      |
| *Dendrobium loddigesii*   | NC_036355        | 152493   | 47.077                      | 94.205                      |
| *Goodyera fumata*         | NC_026773        | 155643   | 48.048                      | 96.151                      |
| *Goodyera procera*        | NC_029363        | 153240   | 47.306                      | 94.667                      |
| *Goodyera schlechtendaliana* | NC_029364 | 154348   | 47.648                      | 95.351                      |
| *Goodyera velutina*       | NC_029365        | 152692   | 47.138                      | 94.328                      |
Table 3 (on next page)

Summary features of chloroplast genome sequences of thirty-four *Dendrobium* species and four *Goodyera* species
Table 3. Summary features of chloroplast genome sequences of thirty-four *Dendrobium* species and four *Goodyera* species

| Organism                        | CDS  | Exon | Gene   | Misc. feature | Repeat region | rRNA | tRNA |
|--------------------------------|------|------|--------|---------------|---------------|------|------|
| *Dendrobium nobile*            | 79   | 22   | 132    | 2             | 2             | 8    | 38   |
| *Dendrobium officinale*        | 76   | 0    | 129    | 0             | 0             | 8    | 38   |
| *Dendrobium strongylanthum*    | 77   | 0    | 130    | 2             | 2             | 8    | 38   |
| *Dendrobium huoshanense*       | 76   | 0    | 129    | 2             | 2             | 8    | 38   |
| *Dendrobium chrysotoxum*       | 63   | 0    | 116    | 2             | 2             | 8    | 38   |
| *Dendrobium nobile* (China)    | 77   | 0    | 130    | 2             | 2             | 8    | 38   |
| *Dendrobium pendulum*          | 76   | 0    | 129    | 2             | 2             | 8    | 38   |
| *Dendrobium moniliforme*       | 73   | 0    | 129    | 11            | 2             | 8    | 39   |
| *Dendrobium primulinum*        | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium aphyllum*          | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium brymerianum*       | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium denneanum*         | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium devonianum*        | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium falconeri*         | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium gratiosissimum*    | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium hercoglossum*      | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium wardianum*         | 71   | 0    | 131    | 16            | 2             | 8    | 38   |
| *Dendrobium wilsonii*          | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium crepidatum*        | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium salaccense*        | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium spatella*          | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium parciflorum*       | 72   | 0    | 131    | 16            | 2             | 7    | 38   |
| *Dendrobium henryi*            | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium chrysanthum*       | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium jenkinsii*         | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium lohohense*         | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium parishii*          | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium ellipsophyllum*    | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| *Dendrobium xichouense*        | 72   | 0    | 132    | 16            | 2             | 8    | 38   |
| Species                        | Rank | Height | Diameter | Leaves | Flowers | Berries |
|-------------------------------|------|--------|----------|--------|---------|---------|
| Dendrobium fimbriatum         | 72   | 0      | 132      | 16     | 2       | 8       | 38      |
| Dendrobium exile              | 72   | 0      | 132      | 16     | 2       | 8       | 38      |
| Dendrobium fanjingshanense     | 72   | 0      | 132      | 16     | 2       | 8       | 38      |
| Dendrobium candidum           | 75   | 0      | 128      | 0      | 0       | 8       | 38      |
| Dendrobium loddigesii         | 68   | 0      | 120      | 9      | 0       | 8       | 39      |
| Goodyera fumata               | 87   | 0      | 133      | 0      | 0       | 8       | 38      |
| Goodyera procera              | 80   | 0      | 127      | 0      | 0       | 8       | 39      |
| Goodyera schlechtendaliana    | 81   | 0      | 129      | 0      | 0       | 8       | 40      |
| Goodyera velutina             | 79   | 0      | 126      | 0      | 0       | 8       | 39      |
Table 4 (on next page)

Counts of nucleotides in the chloroplast genomes
Table 4. Counts of nucleotides in the chloroplast genomes.

| Nucleotide                     | Adenine (A) | Cytosine (C) | Guanine (G) | Thymine (T) | C + G | A + T |
|--------------------------------|-------------|--------------|-------------|-------------|-------|-------|
| Dendrobium nobile              | 46576       | 28853        | 28039       | 48381       | 56892 | 94957 |
| Dendrobium officinale          | 46743       | 28924        | 28107       | 48447       | 57031 | 95190 |
| Dendrobium strongylandium      | 46940       | 29147        | 28431       | 48541       | 57578 | 95481 |
| Dendrobium huoshanense         | 47032       | 29111        | 28316       | 48729       | 57427 | 95761 |
| Dendrobium chrysotoxum         | 47180       | 29400        | 28492       | 48881       | 57892 | 96061 |
| Dendrobium nobile (China)      | 47118       | 28871        | 28748       | 48923       | 57619 | 96041 |
| Dendrobium pendulum            | 46997       | 29122        | 28242       | 48677       | 57364 | 95674 |
| Dendrobium moniliforme         | 45551       | 28339        | 27520       | 47368       | 55859 | 92919 |
| Dendrobium brymerianum         | 46509       | 28968        | 28123       | 48230       | 57091 | 94739 |
| Dendrobium denneanum           | 46440       | 28913        | 28115       | 48097       | 57028 | 94537 |
| Dendrobium devonianum          | 46615       | 28943        | 28108       | 48279       | 57051 | 94894 |
| Dendrobium falconeri           | 46591       | 28911        | 28040       | 48348       | 56951 | 94939 |
| Dendrobium gratiosissimum      | 46521       | 28954        | 28095       | 48259       | 57049 | 94780 |
| Dendrobium hercoglossum        | 46592       | 28941        | 28131       | 48275       | 57072 | 94671 |
| Dendrobium wardianum           | 46479       | 28955        | 28118       | 48236       | 57073 | 94715 |
| Dendrobium wilsonii            | 46668       | 28948        | 28101       | 48363       | 57049 | 95031 |
| Dendrobium crepidatum          | 46482       | 28951        | 28056       | 48228       | 57007 | 94710 |
| Dendrobium salaccense          | 46493       | 28635        | 27735       | 48241       | 56370 | 94734 |
| Dendrobium spatella            | 46524       | 28969        | 28091       | 48245       | 57060 | 94769 |
| Dendrobium parciflorum         | 45941       | 28699        | 27829       | 47604       | 56528 | 93545 |
| Dendrobium henryi              | 46550       | 28936        | 28093       | 48271       | 57029 | 94821 |
| Dendrobium chrysanthum         | 46519       | 28939        | 28078       | 48254       | 57017 | 94773 |
| Dendrobium jenkinsii           | 46497       | 28942        | 28105       | 48173       | 57047 | 94670 |
| Dendrobium lohohense           | 46558       | 28928        | 28098       | 48228       | 57026 | 94786 |
| Dendrobium parishii            | 46487       | 28924        | 28079       | 48199       | 57003 | 94686 |
| Dendrobium ellipsophyllum      | 46690       | 28922        | 28091       | 48323       | 57013 | 95013 |
| Dendrobium xichouense          | 46672       | 28937        | 28098       | 48345       | 57035 | 95017 |
| Dendrobium fimbritatum         | 46483       | 28932        | 28094       | 48164       | 57026 | 94647 |
| Dendrobium exile               | 46251       | 28937        | 28065       | 48041       | 57002 | 94292 |
| Dendrobium fanjingshanense     | 46694       | 28947        | 28115       | 48352       | 57062 | 95046 |
| Dendrobium candidum            | 46695       | 28914        | 28091       | 48394       | 57005 | 95089 |
| Dendrobium fimbriatum          | 46688       | 28934        | 28064       | 48627       | 56998 | 95495 |
| Goodyera fumata                | 48186       | 29569        | 28447       | 49441       | 58016 | 97627 |
| Goodyera procera               | 47095       | 29370        | 28303       | 48472       | 57673 | 95567 |
| Goodyera schlechtendaliana     | 47822       | 29206        | 28146       | 49174       | 57352 | 96996 |
| Goodyera velutina              | 47554       | 28694        | 27658       | 48786       | 56352 | 96340 |
Table 5 (on next page)

Counts of nucleotide frequency in codon positions across the chloroplast genomes
| Nucleotide per position | 1A | 1C | 1G | 1T | 2A | 2C | 2G | 2T | 3A | 3C | 3G | 3T |
|------------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| D. nobile              | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. officinale          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. strongyanthum       | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. huoshanense         | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. chrysotoxum         | 0.3 | 0.19 | 0.28 | 0.22 | 0.29 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. nobile (China)      | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. pendulum            | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. moniliforme         | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. primulinum          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. aphyllum            | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. brymerianum         | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. denneanum           | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. devonianum          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. falconeri           | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. gratoisssimun       | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. hercoglossum        | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. wardianum           | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. wilsonii            | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. crepidatum          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. salaccense          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. spatella            | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. parcellorum         | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. henryi              | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. chrysanthum         | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. jenkinsii           | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. lohohense           | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. parishii            | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. ellipsophyllum      | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. xichouense          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. fimbriatum          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. exile               | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. fanjingshanense     | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. candidum            | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| D. loddigesii          | 0.31 | 0.19 | 0.27 | 0.23 | 0.3 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| G. fumata              | 0.31 | 0.19 | 0.26 | 0.24 | 0.29 | 0.2 | 0.18 | 0.32 | 0.14 | 0.16 | 0.38 |
| G. procera             | 0.31 | 0.19 | 0.26 | 0.24 | 0.29 | 0.2 | 0.17 | 0.33 | 0.32 | 0.14 | 0.16 | 0.38 |
| G. schlechtendaliana   | 0.31 | 0.19 | 0.26 | 0.24 | 0.29 | 0.2 | 0.17 | 0.33 | 0.32 | 0.14 | 0.16 | 0.38 |
| G. velutina            | 0.31 | 0.19 | 0.27 | 0.24 | 0.29 | 0.2 | 0.18 | 0.33 | 0.32 | 0.14 | 0.16 | 0.38 |
Table 6 (on next page)

Relative synonymous codon usage (in parentheses) following the codon frequency across the chloroplast genomes in the genus *Dendrobium*
Table 6. Relative synonymous codon usage (in parentheses) following the codon frequency across the chloroplast genomes in the genus *Dendrobium*.

| Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU |
|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|------|
| UUU(F) | 2018.1 | 1.16 | UCU(S) | 1330 | 1.63 | UAU(Y) | 1371 | 1.38 | UGU(C) | 706.9 | 1.24 |
| UUC(F) | 1459.2 | 0.84 | UCC(S) | 882.8 | 1.08 | UAC(Y) | 621.4 | 0.62 | UGC(C) | 437 | 0.76 |
| UUA(L) | 918.4 | 1.14 | UCA(S) | 999.4 | 1.23 | UAA(*) | 970.5 | 1.05 | UGA(*) | 1065 | 1.15 |
| UUG(L) | 970.9 | 1.21 | UCG(S) | 576.9 | 0.71 | UAG(*) | 732.2 | 0.79 | UGG(W) | 691.4 | 1 |
| CUU(L) | 1069.8 | 1.33 | CCU(P) | 638 | 1.13 | CAU(H) | 919.7 | 1.43 | CGU(R) | 336.1 | 0.63 |
| CUC(L) | 629.2 | 0.78 | CCC(P) | 547.8 | 0.97 | CAC(H) | 369.3 | 0.57 | CGC(R) | 220.7 | 0.41 |
| CUA(L) | 762.8 | 0.95 | CCA(P) | 689.4 | 1.23 | CAA(Q) | 952.8 | 1.38 | CGA(R) | 545.2 | 1.02 |
| CUG(L) | 473.7 | 0.59 | CGG(P) | 375.4 | 0.67 | CAG(Q) | 423.2 | 0.62 | CGG(R) | 343 | 0.64 |
| AUU(I) | 1635.7 | 1.21 | ACU(T) | 646 | 1.21 | AAU(N) | 1580 | 1.39 | AGU(S) | 659.9 | 0.81 |
| AUC(I) | 1072.9 | 0.8 | ACC(T) | 530.8 | 1 | AAC(N) | 695 | 0.61 | AGC(S) | 435.8 | 0.54 |
| AUA(I) | 1337.4 | 0.99 | ACA(T) | 610.3 | 1.15 | AAA(K) | 1914 | 1.31 | AGA(R) | 1171 | 2.2 |
| AUG(M) | 891.4 | 1 | ACG(T) | 343.2 | 0.64 | AAG(K) | 1009 | 0.69 | AGG(R) | 576 | 1.08 |
| GUU(V) | 709.4 | 1.36 | GCU(A) | 467.5 | 1.29 | GAU(D) | 1038 | 1.43 | GGU(G) | 523.7 | 0.99 |
| GUC(V) | 366.7 | 0.7 | GCC(A) | 326.4 | 0.9 | GAC(D) | 413.9 | 0.57 | GGC(G) | 314.4 | 0.59 |
| GUA(V) | 647.8 | 1.24 | GCA(A) | 438.7 | 1.21 | GAA(E) | 1335 | 1.37 | GGA(G) | 754.1 | 1.43 |
| GUG(V) | 366.9 | 0.7 | GCG(A) | 221.5 | 0.61 | GAG(E) | 618.3 | 0.63 | GGG(G) | 521.8 | 0.99 |