The specific DNA barcodes based on chloroplast genes for species identification of Orchidaceae plants

Huili Li1,2, Wenjun Xiao1,2, Tie Tong1, Yongliang Li1, Meng Zhang1, Xiaoxia Lin1, Xiaoxiao Zou1, Qun Wu1 & Xinhong Guo1*

DNA barcoding is currently an effective and widely used tool that enables rapid and accurate identification of plant species. The Orchidaceae is the second largest family of flowering plants, with more than 700 genera and 20,000 species distributed nearly worldwide. The accurate identification of Orchids not only contributes to the safe utilization of these plants, but also it is essential to the protection and utilization of germplasm resources. In this study, the DNA barcoding of 4 chloroplast genes (matK, rbcL, ndhF and ycf1) were used to provide theoretical basis for species identification, germplasm conservation and innovative utilization of orchids. By comparing the nucleotide replacement saturation of the single or combined sequences among the 4 genes, we found that these sequences reached a saturation state and were suitable for phylogenetic relationship analysis. The phylogenetic analyses based on genetic distance indicated that ndhF and ycf1 sequences were competent to identification at genus and species level of orchids in a single gene. In the combined sequences, matK + ycf1 and ndhF + ycf1 were qualified for identification at the genera and species levels, suggesting the potential roles of ndhF, ycf1, matK + ycf1 and ndhF + ycf1 as candidate barcodes for orchids. Based on the SNP sites, candidate genes were used to obtain the specific barcode of orchid plant species and generated the corresponding DNA QR code ID card that could be immediately recognized by electronic devices. This study provides innovative research methods for efficient species identification of orchids. The standardized and accurate barcode information of Orchids is provided for researchers. It lays the foundation for the conservation, evaluation, innovative utilization and protection of Orchidaceae germplasm resources.

Orchidaceae is the second largest family after Composite, and the largest family of monocotyledonous plants1–3. More than 700 genera and more than 20,000 species were identified in the family Orchidaceae, which account for 8 percent of all flowering plants1,2,4–7. Orchids mainly distribute in the tropical and subtropical regions of the world, and a few species grow in the temperate regions2–4,8,9. The Orchidaceae plants exhibit important ornamental, medicinal, research and ecological value2,10–12. Many Orchidaceae plants with beautiful flowers and rich fragrance are ornamental plants, such as Cymbidium, Phalaenopsis, Cypripedium1,3,12–14. Numerous species containing active ingredients, like polysaccharides, alkaloids, phenanthrene and dibenzyl also are served as traditional herbal medicines for treatment of the diseases2,7,10,12,15. These traits that is able to bring great economic benefits make Orchidaceae plants on raising market demand. In the past decades, over-exploitation and habitat destruction by humans caused serious extinction threats to a large number of Orchidaceae plants2,10,15. Additionally, more and more counterfeit and shoddy Orchidaceae-related products emerge. This is not only likely to threaten drug safety, but also caused damage to biodiversity2,7,11,12.

Given that, the accurate identification of Orchidaceae plants is of great significance for their safe utilization, biodiversity and the protection of genetic resources2,7,12,16,17. It is known that traditional identification methods are based on morphological features. Some Orchidaceae plants, however, almost exhibit no morphological differences before flowering, and the morphological features are susceptible to environmental factors2,7,16,18. In addition, there are fewer and fewer experienced experts in morphological identification2,7,12,16,18–20. Totally, this makes the accurate identification to be a time consuming and labor intensive job. Therefore, we are badly in need of a rapid, accessible and accurate identification method.

The DNA barcode technology is a novel molecular recognition technology that uses short and standard DNA fragments for species identification16,17,19,21,22. DNA barcodes were originally utilized to identify

1College of Biology, Hunan University, Changsha 410082, China. 2These authors contributed equally: Huili Li and Wenjun Xiao. *email: gxh@hnu.edu.cn
microorganisms, but now it is able to quickly and accurately identify species at the level of species with unlimited reasons for development stage, internal morphological diversity, factors and users’ professional level. Thus, the DNA barcoding technology has been rapidly applied in species identification, biodiversities, ecological community evolution, species protection, archaeological sample identification, and other aspects. Mitochondrial cytochrome oxidase I gene proposed by Hebert et al. in 2003 had been widely used in animal species identification and phylogenetic development. However, due to the low mutation rate of mitochondrial DNA, mitochondrial cytochrome oxidase I cannot be used in plants. In the past decades, many researchers have made great contributions to the search and application of barcode in plants. Subsequently, many scientists performed a great deal of phylogenetic analyses among numerous families or subfamilies of the orchid family based on two plastid genes matK or rbcL. Many efforts have been made to discover the core barcodes for different land plant taxa, whereas a consensus has not been reached.

After that, CBOL Plant Working Group compared the performance of seven leading candidate plasmon DNA regions (atpF-atpH interval, matK gene, rbcL gene, rpoB gene, rpoC1 gene, psbK-psbI interval and trnH-psbA interval) and recommended the 2-site combination of rbcL+matK as a plant barcode based on the evaluation of recoverability, sequence quality and species identification level. The generality of medicinal plants species identification were assessed according to matK and rbcL genes. The molecular taxonomic identification of the Canarian oceanic hotspot was studied based on matK+rbcL. Chen et al. found that ycf1 showed high identification ability at the species level of rare and protected medicinal plants. The chloroplast gene ndhF was found to be able to identify 100% solanum species by Zhang et al. Although DNA barcoding has been widely studied in phylogeny and species identification of Orchids, it has not been reported that DNA barcoding genes can be used to develop specific identification segments of different species.

Here, we used four chloroplast gene sequences (matK, rbcL, ndhF and ycf1) and three combined sequences including matK+rbcL, matK+ycf1, ndhF+ycf1 of Orchidaceae species to develop unique identification fragments of a certain species of Orchidaceae based on phylogenetic analyses and SNP site analyses. Furthermore, the barcode genes were comprehensively analyzed to obtain standard DNA marker fragments of Orchidaceae. Therefore, this study provided a novel approach, based on the SNP barcode, to accurately and rapidly identify Orchidaceae plants. This technology replenishes traditional methods of identification in Orchidaceae plants. This is the first study to report a strategy for developing specific DNA barcodes of Orchidaceae plants, laying the foundation for the conservation, evaluation, innovative utilization and protection of Orchidaceae germplasm resources.

### Results

#### DNA sequences analysis.

In this study, the sequences including 3040 matK sequences (307 genera, 1900 species), 641 rbcL sequences (55 genera, 192 species), 225 ndhF sequences (102 species, 29 genera), and 384 ycf1 sequences (48 genera, 173 species) of Orchids were obtained from the NCBI Nucleotide database (https://www.ncbi.nlm.nih.gov/) for further analyses.

After blasting and editing, the consensus length of matK, rbcL, ndhF and ycf1 were 2169 bp, 1524 bp, 2953 bp, 8145 bp respectively, and that of combined sequence including matK+rbcL, matK+ycf1, ndhF+ycf1 were 3348 bp, 9731 bp, 9701 bp respectively.

The overall mean nucleotide base frequencies observed for candidate nucleotide sequences and the distribution of the four bases of candidate nucleotide sequences at different coding positions of codons were showed in Table 1. The average number of identical pairs (ii) for candidate nucleotide sequences was showed in Table 2. The account of transitional pairs (si) and transversional pairs (sv) of nucleotide sequences was showed in Table 2. The transitional and transversional of bases in the sequences may be related to the species difference.

Polyorphism site analysis of the candidate nucleotide sequences revealed in Table 3. Among the single sequence and the combination sequence rbcL sequence had the least proportion of mutation sites, accounting for 34.8%, while the conservative sites in the corresponding rbcL sequence accounted for 64.7%. The sequence matK had the highest proportion of mutation sites (70.2%), and the corresponding matK sequence had the lowest proportion of conservative sites (18.9%).

#### Genetic diversity.

There must be some genetic variation based on their species differences since the data used to analyze were obtained from different species. The basic indicators of genetic diversity, displayed in Table 4, worked out in accordance with pairwise nucleotide differences and nucleotide diversity, and the validity of these indexes were verified by two neutrality tests, like Fu’s F and Tajima’s D. The matK+ycf1 sequences had revealed maximum genetic diversity cumulatively on the base of Eta value, revealed 2314 mutations within all sequences. While the rbcL sequences only had 322 mutations variations in all sequences. The significance of genetic diversity was verified by both neutrality tests, which confirmed that all sequences had significant differ-

| Sequences | Base content | A | T | C | G | GC | AT | GC | AT | GC |
|-----------|--------------|---|---|---|---|----|----|----|----|----|
| matK      |              | 30.8 | 37.8 | 16.4 | 15.0 | 31.4 | 68.5 | 31.5 | 68.9 | 31.2 | 68.3 | 31.7 |
| rbcL      |              | 28.0 | 39.2 | 18.4 | 24.5 | 42.9 | 62.7 | 37.3 | 53.2 | 46.7 | 55.4 | 44.6 |
| ndhF      |              | 27.3 | 39.4 | 16.1 | 17.2 | 33.3 | 66.9 | 33.1 | 64.9 | 35.1 | 68.4 | 31.6 |
| ycf1      |              | 40.4 | 50.0 | 13.9 | 15.7 | 29.6 | 69.3 | 30.8 | 71.8 | 28.2 | 60.2 | 29.8 |

Table 1. The nucleotide base frequencies analysis of candidate nucleotide sequences in Orchidaceae plants.
ence but no very significant difference based on the probability value (p-value) of Fu's Fs test and Tajima's D test (Table 4).

Like the neutrality testes of the Tajima test statistic (D value) in the sequences, the genetic variation for ndhF sequences was negatively little higher (−2.37565) with respect to rbcL sequence, consisting value up to −0.51015. And for combined sequences, the genetic variation for ndhF + ycf1 sequences was negatively little higher (−2.13392) with respect to matK + rbcL sequence, consisting value up to −1.75132. With respect to Fu's Fs value for sequences variation, the ndhF sequences was higher sequences variation (−2.96843), shown in Table 4, in comparison with rbcL sequence (−0.51015). In order to observe nucleotide mismatch distribution among different sequences of Orchidaceae species, DNA sequences were analyzed for population size changes which was enriched the results of genetic diversity among species. All results showed significant genetic variation in Orchidaceae species for candidate nucleotide sequences (Fig. 1).

**Phylogenetic analysis.** In this study, we used the MEGA7.0 software based on the Neighbor Joining method and Kimura 2-parameter model to identify rbcL, ndhF and ycf1 sequence of the evolutionary tree, and we compressed the same genera or the same subtribes of Orchid with the MEGA 7.0 own Compress Subtree. In

### Table 2. The analysis of nucleotide pair frequencies of candidate nucleotide sequences of Orchidaceae plants. ii Identical Pairs, si Transitionals Pairs, sv Transversionsal Pairs, R si/sv.

| Sequence | ii Avg | 1st | 2nd | 3rd | si Avg | 1st | 2nd | 3rd | sv Avg | 1st | 2nd | 3rd | R si/sv |
|----------|-------|-----|-----|-----|--------|-----|-----|-----|--------|-----|-----|-----|---------|
| matK     | 1268  | 431 | 423 | 414 | 49     | 16  | 15  | 18  | 43     | 13  | 15  | 15  | 1.1     |
| rbcL     | 1378  | 462 | 458 | 458 | 29     | 10  | 10  | 9   | 14     | 4   | 5   | 5   | 2.1     |
| ndhF     | 1234  | 410 | 419 | 404 | 38     | 14  | 12  | 12  | 34     | 11  | 10  | 12  | 1.1     |
| ycf1     | 4521  | 1514| 1509| 1498| 205    | 61  | 72  | 72  | 200    | 68  | 62  | 70  | 1.0     |
| matK + rbcL | 2835 | 953 | 936 | 946 | 73     | 19  | 32  | 22  | 56     | 15  | 21  | 21  | 1.3     |
| matK + ycf1 | 6015 | 1990| 2008| 2017| 247    | 94  | 76  | 77  | 239    | 82  | 81  | 75  | 1.0     |
| ndhF + ycf1 | 5718 | 1905| 1914| 1899| 189    | 63  | 61  | 65  | 187    | 61  | 61  | 65  | 1.0     |

### Table 3. The analysis of variation of candidate barcode sequences in Orchidaceae plants.

| Sequences | n  | Nucleotide diversity | π   | Neutrality tests |
|-----------|----|----------------------|-----|-----------------|
|           | S  | k                    | θ   | Fu's Fs         |
|           |    |                      |     | p-value         |
|           |    |                      |     | D               |
|           |    |                      |     | p-value         |
| matK     | 3050 | 3.12596 | 1523 | 0.9050 | 0.24339 | 0.07270 | −2.57476 | <0.05 | −1.84286 | <0.05 |
| rbcL     | 643  | 3.12596 | 1523 | 0.9779 | 0.07155 | 0.02503 | −0.51015 | <0.10 | −1.92689 | <0.05 |
| ndhF     | 234  | 3.12596 | 1523 | 0.9660 | 0.18546 | 0.04589 | −2.96843 | <0.05 | −2.37565 | <0.01 |
| ycf1     | 384  | 3.12596 | 1523 | 0.9921 | 0.17379 | 0.07339 | 0.25100 | <0.10 | −1.78584 | <0.05 |
| matK + rbcL | 372  | 3.12596 | 1523 | 0.9924 | 0.12616 | 0.05462 | −0.11121 | <0.10 | −1.75132 | <0.05 |
| matK + ndhF | 216  | 3.12596 | 1523 | 0.9853 | 0.12276 | 0.04604 | −1.32301 | <0.10 | −2.00313 | <0.05 |
| matK + ycf1 | 378  | 3.12596 | 1523 | 0.9932 | 0.15463 | 0.06542 | 0.08057 | <0.10 | −1.79023 | <0.05 |
| ndhF + ycf1 | 228  | 3.12596 | 1523 | 0.9740 | 0.15585 | 0.05191 | −1.47948 | <0.10 | −2.13392 | <0.01 |
Figure 1. Pairwise mismatch distributions, based on matK, rbcL, ndhF, ycf1 and the combined sequences by DnaSP v5. Note: The X-axis shows the observed distribution of pairwise genetic variation, and the Y-axis shows the frequency. $R^2$ Ramos-Ornsins and Rozas statistics, $R$ Raggedness statistic, $\tau$ Date of the Growth or Decline measured of mutational time, C.V. Coefficient of variation.
Figure 2. The NJ tree of Orchidaceae coming from analysis of the cp DNA matK sequence based on the K2P model. Names tagged in red indicates the genus, tagged in green showed the subtribe and tagged in blue showed the subfamily; The Numbers on the branches represent more than or equal to 50 percent support after the 1000 bootstrap replications test; Numbers following taxon names showed the number of species.
Figure 3. The NJ tree of Orchidaceae coming from analysis of the cp DNA rbcL sequence based on the K2P model. Names tagged in red indicates the genus and tagged in green showed the subtribe; The Numbers on the branches represent more than or equal to 50 percent support after the 1000 bootstrap replications test; Numbers following taxon names showed the number of species.
the light of the topological structure of the evolutionary tree, species in several subfamilies are not be well identified based on matK, rbcL and the combined sequence matK + rbcL. In contrast, the ndhF, ycf1 sequences and the combined sequences matK + ycf1 and ndhF + ycf1 of chloroplast genes exhibit better identification ability at the generic level (Figs. 2, 3, 4, 5, 6, 7, 8).

**Analysis of barcoding gap.** An ideal DNA barcoding sequence for species identification should satisfy that inter-specific genetic variation is significantly greater than intra-specific genetic variation. In order to more accurately assess individual chloroplast genes and combined sequences in the Orchid genus species, and to verify the applicability of candidate sequences, the barcoding gap was analyzed according to frequency distribution showed in Fig. 9. The results revealed that the ndhF gene showed better performance in a single gene, while the combined sequences of ndhF + ycf1 showed the best performance. The results of the Best Close Match of several candidate barcodes based on genetic distance are showed in Table 5. Among the single genes, the accuracy rate of ycf1 gene for orchid plant identification is 89.32%, with 3.38% fuzzy identification rate and 6.25% error identification rate. The ndhF gene exhibits the highest identification rate and lower error rate of matK + ycf1, followed by ndhF + ycf1 sequence. The accuracy of matK + ycf1 sequence was 89.6%, with 2.8% fuzzy identification rate and 1.12% error identification rate. The accuracy rate of ndhF + ycf1 sequence was 88.78%, with 2.33% fuzzy identification rate and 2.8% error identification rate. The data indicated that ndhF and ycf1 were suitable for

**Figure 4.** The NJ tree of Orchidaceae coming from analysis of the cp DNA ndhF sequence based on the K2P model. Names tagged in red indicates the genus and tagged in green showed the subtribe; The Numbers on the branches represent more than or equal to 50 percent support after the 1000 bootstrap replications test; Numbers following taxon names showed the number of species.
Figure 5. The NJ tree of Orchidaceae coming from analysis of the cp DNA ycf1 sequence based on the K2P model. Names tagged in red indicates the genus and tagged in green showed the subtribe; The Numbers on the branches represent more than or equal to 50 percent support after the 1000 bootstrap replications test; Numbers following taxon names showed the number of species.
Figure 6. The NJ tree of Orchidaceae coming from analysis of the matK + rbcL sequence based on the K2P model. Names tagged in red indicates the genus and tagged in green showed the subtribe. The Numbers on the branches represent more than or equal to 50 percent support after the 1000 bootstrap replications test. The Numbers following taxon names showed the number of species.
**Figure 7.** The NJ tree of *Orchidaceae* from analysis of the matK + ycf1 sequence based on the K2P model. Names tagged in red indicates the genus and tagged in green showed the subtribe; The Numbers on the branches represent more than or equal to 50 percent support after the 1000 bootstrap replications test; Numbers following taxon names showed the number of species.
Figure 8. The NJ tree of Orchidaceae from analysis of the ndhF + ycf1 sequence based on the K2P model. Names tagged in red indicates the genus and tagged in green showed the subtribe.

Figure 9. Histogram of frequency of intra-species (black) and inter-species (red) of Orchidaceae based on K2P distance of candidate genes. The X-axis represents the genetic distance, and the Y-axis represents the frequency.
the identification of Orchids at the level of genus and species, while the combined sequences of \( \text{matK} + \text{ycf1} \) and \( \text{ndhF} + \text{ycf1} \) were qualified at the genera and species levels.

Specific barcodes based on SNP sites. Based on SNP sites, species-specific barcodes were developed and the appropriate fragments were blasted into the NCBI database. Based on the \( \text{ndhF} \) sequence, the specific barcode of species \( \text{Dendrobium scoriarum} \) was obtained. Knowledge about specific barcodes of species \( \text{Neuwiedia thelymitra} \), \( \text{Spiranthes sinensis} \) and \( \text{Epiblema cocflorum} \) based on \( \text{ycf1} \) sequence was obtained. Based on the combined sequence \( \text{ndhF} + \text{ycf1} \), the specific barcodes of \( \text{Liparis loeselii} \), \( \text{Cremastraa appendiculata} \), \( \text{Spiranthes sinensis} \) and \( \text{Anathallis obovata} \) were obtained, whereas \( \text{Liparis loeselii} \) and \( \text{Cremastra appendiculata} \) had two specific barcodes. Two-Dimensional code can be scanned by electronic equipment from DNA fragments that can be used for species identification. It can provide theoretical support for subsequent researchers. Using the Two-Dimensional code coding method, the species-specific barcode obtained was converted into two-dimensional barcode image, which was conducive to the conversion of barcode information (Figs. 10, 11).

**Discussion**

DNA barcode is able to be utilized for species identification by means of a DNA fragment that is common to all species. The fragment must simultaneously contain adequate variability to allow for species identification and enough conservative area for the design of universal primers\(^2^1,2^3\). So far, DNA barcoding have been widely used in many genera of \( \text{Orchidaceae} \)\(^1,2,7,16\). As far as we know, it is the first time that multi-aspect analysis in species identification of \( \text{Orchidaceae} \) with such a well-rounded species size, based on \( \text{matK} \) and \( \text{rbcL} \) regions.

The results of sequences analyses on average GC content showed that the GC content of candidate sequences of Orchids was far less than AT content, while significantly less than the GC content of about 50% in common angiosperms. Of sequence variation situation analysis, the candidate gene mutations exist base insert and missing phenomenon. We performed the analysis of the genetic diversity by the DnasP 5.0 software. The higher haploid type diversity and relatively low haploid type diversity of nucleotide diversity demonstrated that the candidate sequences had certain polymorphisms.

The CBOL recommends \( \text{matK} \) and \( \text{rbcL} \) as universal barcodes in plant kingdom\(^2^3\). With the development of science and technology, many subsequent scientists have evaluated the discriminability of different DNA barcoding genes in different families or genera, but the discriminability of a candidate gene in different plants was different.

On the basis of phylogenetic relationship, the Barcoding Gap and the Best Close Match with the genetic distance in evaluating candidate barcode identification capability in Orchid, the phylogenetic analyses showed that the identification ability of \( \text{matK} \) and \( \text{rbcL} \) was low on the genus level. The possible reason was that there were more species in this study, which made the species in the related genus unable to form branches alone. The sequences of \( \text{ndhF} \) and \( \text{ycf1} \) were suitable for identification of genus and species of Orchids, and the combined sequences \( \text{matK} + \text{ycf1} \) and \( \text{ndhF} + \text{ycf1} \) were qualified at the genera and species levels. The Barcoding Gap test indicated that these candidate genes all contained Barcoding Gap, and the variation between species and within

| Sequences      | Correct | Fuzzy | Error | Did not identify |
|----------------|---------|-------|-------|------------------|
| \( \text{ndhF} \) | 88.65%  | 3.45% | 2.50% | 5.48%            |
| \( \text{ycf1} \)  | 89.32%  | 3.38% | 6.25% | 1.05%            |
| \( \text{matK} + \text{ycf1} \) | 89.60%  | 2.80% | 1.12% | 6.48%            |
| \( \text{ndhF} + \text{ycf1} \) | 88.78%  | 2.33% | 2.80% | 6.09%            |

Table 5. Best Close Match test results based on genetic distance.

![Figure 10](https://example.com/figure10.png)

**Figure 10.** DNA barcodes and two-dimensional DNA barcodes of Orchidaceae species based on \( \text{ndhF} \) and \( \text{ycf1} \) genes. Base A in green, base T in red, base C in blue, and base G in black.
species had clear boundaries. The test results of Best Close Match revealed that the all combined sequences exhibited high genus identification rate, which was suitable for the identification of orchids at the level of genus and species.

Based on the SNP sites, the species level specific DNA barcodes of *Orchid* were successfully developed. Combinatorial sequences were able to develop more species-specific barcodes than chloroplast genes, which might be the result of combination sequences could provide more mutation sites and SNP sites. There were some differences in the specificities of different combination genes in Orchidaceae plants. Compared with *ndhF* + *ycf1*, the combined sequences of *matK* + *ycf1* could be developed more specific barcodes, which might be related to the species identification accuracy of *matK* + *ycf1* in Orchids.

**Conclusion**
In summary, *ndhF*, *ycf1*, *matK* + *ycf1* and *ndhF* + *ycf1* sequences are competent to develop species-specific barcodes to identify Orchidaceae plants at the molecular level. Cluster analysis using the *ndhF*, *ycf1*, *matK* + *ycf1* and *ndhF* + *ycf1* sequences in Orchid are nearly consistent with traditional plant morphology. Additionally, this study not only broadens the application of the *matK* and *rbcL* sequences in the barcode field, but also provides a novel thought to expand species identification method in a wide range of plant at the species level.

**Figure 11.** DNA barcodes and two-dimensional DNA barcodes of Orchidaceae species based on *matK* + *ycf1* and *ndhF* + *ycf1* genes. Base A in green, base T in red, base C in blue, and base G in black.
**Methods**

**Nucleotide sequences.** For species identification, we retrieved the chloroplast DNA reference sequences including *matK, rbcL, ndhF* and *ycf1* from the NCBI Gene database (https://www.ncbi.nlm.nih.gov/). We obtained the combined sequence including *matK + rbcL, matK + ycf1, ndhF + ycf1* by supermat’s function in R Phylotools package. After manual screening, the short nucleotide sequences were deleted, and the sequences with different directions were modified manually.

**Data analysis.** We performed the sequences alignment by the Muscle in the MEGA 7.0 software (https://www.megasoftware.net/) with the default alignment parameters for multiple sequences alignment parameters. In the pairwise distances analyses, the positions containing gaps and missing were eliminated from the data set (complete deletion option). Phylogenetic trees constructed with the Neighbor-joining (NJ) method according to Kimura 2-Parameter (K2P) model was assessed by the MEGA 7.0 (S2). The clade reliability in these trees using the NJ methods was tested by bootstrapping, in which 1000 repeated sampling tests were performed to obtain the support values of the clade nodes. Polymorphic site, genetic diversity indices and neutrality tests [Fu’s Fs (8) and Tajima’s D (3)] were performed by the DnaSP v5 (http://www.ub.edu/dnasp/index_v5.html).

Received: 17 December 2019; Accepted: 4 January 2021
Published online: 14 January 2021

**References**

1. Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M. & André, S. An updated classification of Orchidaceae. *Bot. J. Linn Soc.* 177, 151–174 (2015).

2. Kim, H. M., Oh, S. H., Bhandari, G. S., Kim, C. S. & Park, C. W. DNA barcoding of *Orchidaceae* in Korea. *Mol. Ecol. Resour.* 14, 499–507 (2014).

3. Gamara, R., Cela, P. G. & Ortúñez, E. *Orchidaceae* in equatorial guinea (west tropical Africa): Nomenclatural and taxonomic notes, new records and critical taxa. *New Bull.* https://doi.org/10.1007/s11225-018-9787-9 (2019).

4. Atwood, J. T. The size of the *Orchidaceae* and the systematic distribution of epiphytic Orchids. *Selbyana* 9, 171–186 (1986).

5. Arditti, J. *Fundamentals of Orchid Biology*, Vol. 691 ( Wiley, New York, 1992).

6. Douzery, E. J. P., Pridgeon, A. M., Kores, P., Linder, H. P. & Chase, K. M. W. Molecular phylogenetics of disease (*Orchidaceae*): A contribution from nuclear ribosomal ITS sequences. *Am. J. Bot.* 86, 887–899 (1999).

7. Feng, S. G. et al. Molecular identification of *Dendrobium* species (*Orchidaceae*) based on the DNA barcode ITS2 region and its application for phylogenetic study. *Int. J. Mol. Sci.* 16, 21975–21988 (2015).

8. Schuiteman, A. Devogela (*Orchidaceae*): A new genus from the moluccas and new guinea. *Bhumea* 49, 361–366 (2004).

9. Nino, D. M., Plunkett, G. M. & Driossart, V. New phylogenetic insights toward developing a natural generic classification of African angraecoid *Orchidaceae* (*Vandeae Orchidaceae*). *Mol. Phylogenet. Evol.* 126, 241–249 (2014).

10. Pérez, G. & Rosa, M. Orchids: A review of uses in traditional medicine, its phytochemistry and pharmacology. *J. Med. Plants Res.* 4, 592–638 (2010).

11. Jacqueyn, H., Merckx, V. & Brys, R. Analysis of network architecture reveals phylogenetic constraints on mycorrhizal specificity in the genus *Orchis* (*Orchidaceae*). *New Phytol.* 192, 518–528 (2011).

12. Tri, S. P. & Atwood, J. T. The size of the *Orchidaceae* and the systematic distribution of epiphytic Orchids. *Selbyana* 9, 171–186 (1986).

13. Yoshikawa, M., Murakami, T. & Kishi, A. Novel indole S, O-bisdesmoside, calanthoside, the precursor glycoside of tryptanthrin, indrubin, and isatin, with increasing skin blood flow promoting effects, from two Calanthe species (*Orchidaceae*). *Chem. Pharm. Bull.* 46, 886–888 (1998).

14. Watanabe, K., Tanaka, R. & Sakurai, H. Structure of cymbidine A, a monomeric peptidoglycan-related compound with hypotensive and diuretic activities, isolated from a higher plant, *Cymbidium goeringii* (*Orchidaceae*). *Chem. Pharm. Bull.* 55, 780–783 (2007).

15. Eda, K. & Budak, K. B. Detection and quantification of salep with real time PCR utilizing the nr-its2 region. *PLoS ONE* e0115168 (2015).

16. Asahina, H., Shinozaki, J. & Masuda, K. Identification of medicinal *Dendrobium* species by phylogenetic analyses using *matK* and *rbcL* sequences. *J. Natl. Med. 64, 133–138 (2010).

17. De Boer, H. J. et al. DNA metabarcoding of orchid derived products reveals widespread illegal orchid trade. *Proc. R. Soc. Lond. B Biol. Sci.* 284, 1863 (2017).

18. Techne, N., Parveen, I., Pan, Z. & Khan, I. A. DNA metabarcoding of medicinal plant material for identification. *Curr. Opin. Biotech.* 25, 103–110 (2014).

19. Devos, N., Oh, S. H., Raspe, O., Jacquemart, A. L. & Manos, P. S. Nuclear ribosomal DNA sequence variation and evolution of spotted marsh orchids *Orchids* (*Dactylorhiza maculata group*). *Mol. Phylogenet. Evol.* 36, 568–580 (2005).

20. Zhang, G. Q., Liu, K. W., Li, Z., Lohaus, R. & Hsiao, Y. Y. The *Apostasia* genome and the evolution of orchids. *Nature* 549, 379–383 (2017).

21. Hebert, P. & Gregory, T. R. The promise of DNA barcoding for taxonomy. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12794–12797 (2009).

22. Coissac, E., Hollingsworth, P. M. & Lavergne, S. From barcodes to genomes: Extending the concept of DNA barcoding. *Mol. Ecol.* 25, 423–428 (2016).

23. CBOL Plant Working Group. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12794–12797 (2009).

24. Cameron, K. M., Chase, M. W. & Whitten, W. M. A phylogenetic analysis of the *Orchidaceae*: evidence from *rbcL* nucleotide sequences. *Am. J. Bot.* 86, 208–224 (1999).

25. Lahaye, R. & Van der Bank, M., Bogar, D., Warner, J., Pupulin, F., DNA barcoding the florals of biodiversity hotspots. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2923–2928 (2008).

26. Ma, H. L., Zhu, Z. B. & Zhang, X. M. Species identification of the medicinal plant *Tulipa edulis* (*Liliaceae*) by DNA barcode marker. *Biochem. Syst. Ecol.* 35, 362–368 (2014).

27. Xu, S. et al. Evaluation of the DNA barcodes in *Dendrobium* (*Orchidaceae*) from mainland Asia. *PLOS ONE* e0115168 (2015).

28. Hebert, P. D. N., Ratnasingham, S. & De, W. J. R. Barcoding, *Dendrobium* and the evolution of orchids. *Nature* 549, 379–383 (2017).

29. Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. Identification of birds through DNA barcodes. *PLoS Biol.* e312 (2004).

30. Salazar, G. A., Chase, M. W. & Ingrouille, A. M. Phylogenetics of *Cranichideae* with emphasis on *Spiranthinae* (*Orchidaceae*, *Cranichideae*): Evidence from plastid and nuclear DNA sequences. *Am. J. Bot.* 90, 777–795 (2003).

31. Kress, W. J. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369–8374 (2005).
