Systematics, epidermal defense and bioprospecting of wild orchids

Kusuma Wati, R.

Chapter 3

Identification of sterile orchids in living collections by DNA barcoding of types

Richa Kusuma Wati, Esmée de Graaf, Erik Smets, Rogier R. van Vugt, Aninda Retno Utami Wibowo, Muhammad Bima Atmaja, Barbara Gravendeel

23rd Proceedings of the World Orchid Conference, Taiwan, April 2021

Abstract. Orchid collections in botanic gardens are an essential source for biodiversity research. Most species require very specific temperature, humidity, light levels and nutrient concentrations for flower induction and survival and therefore often remain sterile or die shortly after collection from the wild. This severely hampers the identification of such collections to species level. DNA barcodes obtained from fertile type specimens in herbaria are a potential tool for fast species identification of sterile living collections, but only a few curators of herbaria allow destructive sampling of type specimens. We obtained permission to perform destructive DNA extraction of a small number of the numerous leaves from a total of 32 type and 11 non-type specimens of the poorly known necklace orchid genus *Glomera* Blume preserved in the herbarium of Naturalis Biodiversity Center (L) and Herbarium Bogoriense (BO) that were collected between 5 up to 194 years ago. We used four primer combinations to fully sequence the nrITS region of these specimens and fresh sterile living collections and obtained Sanger sequences for 38 dried specimens and 43 living collections with a length varying between 70-887 bp. A possible correlation between age and length of DNA barcodes retrieved was investigated. With the short sequences obtained, a total of 6 sterile living collections could be identified to species level. No correlation between age and relative length of DNA barcodes retrieved was found. A total of 38 living collections remained unnamed and are possibly new to science. Our result shows that DNA barcodes obtained from type material can provide reliable taxonomic information of sterile living collections. We propose a less rigorous policy regarding permission to generate DNA barcodes from type specimens to improve the identification of sterile specimens in living collections for better protection of poorly known orchid genera.

3.1 Introduction

Many countries have a different strategy for plant conservation that can either take place in situ, ex situ or both, and at species and/or population level. Plant conservation supports the recovery and reintroduction of endangered species by developing a gene bank and/or restoration ecology (Heywood, 2017). A botanic garden is the equivalent of a museum of living plants where species can be studied, not only by experts to increase scientific knowledge, but also by general plant enthusiasts for educational and recreational purposes (Holttum, 1970). Orchidaceae are one of the most diverse angiosperm families with more than 25,000 species (Dressler, 1993). Because of their spectacular diversity in life forms and flowers, orchids are usually present in collections of botanic gardens. For conservation purposes, many orchids have been collected from different locations and habitats. These orchid collections have been important sources for studies on conservation, ecology, evolution, germination, physiology, pollination and taxonomy. Many new orchid taxa have been described from specimens in collections of botanic gardens. Especially tropical orchids have highly diverse flowers that evolved due to highly specific plant-pollinator interactions (Kim et al., 2014). In greenhouses, these collections are kept alive, but due to the highly diverse requirements of temperature, light levels and nutrition for each species, it is a challenge to stimulate flowering (Wang, 2000; Lopez and Runkle, 2004, 2005; Vaz et al., 2004; Pfeifer et al., 2006). Some accessions therefore remain sterile for their entire life, or die shortly after collection from the wild. Since it is extremely difficult or impossible to identify tropical orchids lacking flowers (Cameron et al., 1999), many specimens in living collections therefore remain unnamed.

Accurate identification of newly discovered species and a formal description are essential for plant conservation (Lahaye et al., 2008; Gutiérrez, 2010; Strutzenberger et al., 2012). Due to extensive variation of morphological characters within a single species, descriptions of new species solely based on morphological characters are not published that much anymore (Bogarín et al., 2018) especially when it comes down to the lesser-known orchid genera. DNA barcoding is an established tool for rapid identification of a species with a short

standard DNA sequence (Hebert et al., 2003). Fresh tissue is usually harvested to obtain DNA sequences of high quality.

When fresh material is not available, dried material may be tried. Several studies showed the use of herbarium material as an underutilised genomic treasure (Blattner, 1999; Eloff, 1999; Särkinen et al., 2012; Bakker et al., 2016; Xu et al., 2015; Hart et al., 2016). Especially for type specimens, most herbarium curators do not allow destructive sampling, even if only a small part of the type is requested. Arguments used are that destructive sampling for DNA extraction might cause irreparable damage to the specimens, which conflicts with their historical and scientific importance (Staats et al., 2011).

In this study, we focused on *Glomera* Blume, an underexplored genus from subtribe Coelogyninae. *Glomera* currently comprises 169 species that are widely distributed in Indonesia, Papua New Guinea, The Philippines, Bismarck Archipelago, Solomon, New Caledonia, and Fiji (Wati, van Vugt and Gravendeel, 2018). We studied the living *Glomera* collections in the greenhouse of the Hortus botanicus Leiden and Bali Botanic Gardens, where many specimens remained unidentified because they never flowered in cultivation. Since these orchids have many small leaflets, we obtained permission from the curators of the herbaria of Naturalis Biodiversity Center in the Netherlands and Herbarium Bogoriense in Indonesia to generate DNA barcodes of type specimens to identify sterile orchids in collection of the Hortus botanicus and Bali Botanic Gardens.

3.2 Material and Methods

3.2.1 Taxon sampling

We sampled a total of 81 *Glomera* specimens. A total of 43 fresh leaf fragments were collected from living plants in the orchid greenhouses of the Hortus botanicus Leiden, The Netherlands, and Bali Botanic Gardens, Indonesia. A total of 38 leaf samples were collected from dried herbarium specimens present in the herbaria of Leiden (L) and Bogor (BO), where most types of the genus *Glomera* are deposited. Of the 38 herbarium samples analysed, 24 were type specimens, belonging to species collected in the same geographical regions as the orchids from the living collections. Only a small piece of leaf tissue (1 cm2) was collected. The specimen age ranged from 5 to 194 years (see Table 3.1).

Table 3.1. Details of herbarium specimens analyzed in this study.

Species	Specimen voucher	Specimen collection year
G. acutiflora (Schltr.) J.J.Sm.	L0061314	1906
G. acutiflora (Schltr.) J.J.Sm.	L1521087	2003
G. amboinensis (Ridl.) J.J.Sm.	L0043490	1912
G. diosmoides (Schltr.) J.J.Sm.	L0061263	1965
G. angiensis J.J.Sm.	LSJ 1861	2013
G. bambusiformis Schltr.	L0061317	1908
G. compressa J.J.Sm.	L0056417	1920
G. confusa J.J.Sm.	RK 197	2013
G. erythrosma Blume	L0061325	1825
G. gamosepalata P.Royen	L0063654	1938
G. hamadryas (Schltr.) J.J.Sm.	L0056420	1953
G. hamadryas (Schltr.) J.J.Sm.	L0056422	1913
G. hamadryas (Schltr.) J.J.Sm.	L0056421	1910
G. hamadryas (Schltr.) J.J.Sm.	L0056423	1908
G. hamadryas (Schltr.) J.J.Sm.	L0064472	1908
G. hamadryas (Schltr.) J.J.Sm.	L0056419	1920
G. kanke P.Royen	L0056414	1971
G. kanke P.Royen	L0061331	1961
G. kanke P.Royen	L0064500	1972
G. noroma (P.Royen) J.M.H.Shaw	L 0061278	1961
G. palustris J.J.Sm.	L0056415	1965
G. palustris J.J.Sm.	L0061334	1920
G. pullei J.J.Sm.	L0426043	1913
G. secunda J.J.Sm.	L0043484	1918
G. sp.	LSJ 1885	2016
G. sp.	LSJ 1850	2016
G. sp.	LSJ 1862	2016
G. sp.	Droissart & Juswara 1779	2014
G. sp.	RK 126	2013
G. sp.	RK 18	2013

G. sp.	WB 483	2014
G. sp.	GT 2383	2006
G. sp.	WB 476	2014
G. sp.	LSJ 1854	2016
G. sp.	GT 2894	2010
G. sp.	LSJ 1859	2016
G. sp.	HBL 20111281	2011
G. viridis (Schltr.) J.J.Sm.	LSJ 1882	2016

3.2.2 DNA extraction, amplification and sequencing

Leaf fragments were placed in 2 ml reaction tubes containing a glass bead (7 mm diameter) and homogenized in a Retsch Mixer Mill, MM400 for 3 minutes at 30 Hz/second into a fine powder. Total genomic DNA of herbarium and silica-dried material was extracted using the 2x CTAB (Cetyltrimethylammonium bromide) method (Doyle and Doyle, 1987) with longer DNA precipitation steps for the herbarium samples: these extracts were stored at -20°C for one week to obtain a higher DNA yield. The nuclear ribosomal ITS-5.8S-ITS2 (nrITS) region of silica-gel dried of freshly harvested leaf material was amplified using primers 17SE (5'-ACGAATTCATGGTCCGGTGAAGTGTTC-3') and 26SE (5'-TAGAATTCCCCGGTTCGCTCGCCGTTAC-3') as described by Sulistyo et al. (2015).

Subsequently, a M13 universal sequencing primer was added to the 5' end of the forward (ACGAATTCATGGTCCGGTGAAGTGTTC) and reverse (TA-GAATTCCCCGGTTCGCTCGCCGTTAC) primers to improve Sanger sequencing efficiency. Each PCR reaction had an end volume of 25 µl, containing the template DNA, CoralLoad PCR buffer (Qiagen), dNTPs, Taq DNA Polymerase (Qiagen), and both primers. The PCR reactions were carried out using a C1000 Touch Thermal Cycler (Bio-Rad). The following thermocycling program was used to amplify each gene fragment: 96 °C for 5 min (1 cycle); 96 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min (35 cycle); and then 72 °C for 7 min (1 cycle).

The nrITS region of herbarium preserved leaf material was amplified using primer p3 (5'-GACTCYCGGCAATGGATATCTCG-3') and p4 (5'-CCGCTTATTGATATGCTTAAACTCRGC-3') as described by Cheng et

al. (2016) and primer F1 (5'-CGAGTCTTTGAACGCAAGTTGCG-3') and R1 (5'-GGCCAACGAGACGATAACCC-3'), F2 (5'-CTGCGGAAGGATCAT TGTCGAGAC-3') and R2 (5'-TCAAACCGGCGCAGCTTCG-3'), R3 (5'-CGGCTCTCGCATCGATGAAGAG-3'), F3 (5'-ACCAACAGCAAGG-3') and R4 (5'-GCTTGGAATGCGACCCCAGG-3') that were newly designed for this study. Each PCR reaction had an end volume of 25 µl, containing the template DNA, 5x Phire PCR buffer (ThermoScientific), BSA, dNTPs, Phire Hot Start II DNA Polymerase (ThermoScientific), and both primers. The following thermocycling program was used to amplify each gene fragment: 98 °C for 1 min (1 cycle); 98 °C for 10 s, 50 °C for 10 s, 72 °C for 20 s (40 cycle); and then 72 °C for 1 min (1 cycle).

Sanger sequences were assembled and edited in Geneious ® R8 (Biomatters Ltd., Auckland, New Zealand) (Kearse et al., 2012). The ends of all data sets were trimmed to avoid character misinterpretation. Ambiguous bases were replaced with "N" in the data matrix. DNA sequences were aligned using the MAFFT platform (Multiple Alignment Fast Fourier Transform (Katoh and Standley, 2013) as implemented in Geneious ® R8 with subsequent manual adjustment. Missing data were replaced with "?".

3.2.3 Phylogenetic analysis

To assess whether sequences from nrITS formed species-specific clusters, we conducted distance-analysis using Neighbour-Joining (NJ) analyses and phylogenetic analysis using Maximum Likelihood (ML), Maximum Parsimony, and Bayesian interference (BI) with *Coelogyne fimbriata* Lindl. as outgroup based on earlier studies (Gravendeel et al., 2001) that showed this genus most closely related to the genus *Glomera*. The chosen nucleotide substitution model GTR+G was calculated using the Akaike Information Criterion (AIC) in jModelTest2 v.2.1.6 (Darriba et al., 2015). The analyses were run in the CIPRES Science Gateway v.3.1. (Miller et al., 2010). We performed Bayesian interference analyses with Mr.Bayes v.3.2.6 on XSEDE (Huelsenbeck et al., 2004) with the following parameters: number of runs (nruns=2), number of chains to run (nchains=4), number of generations (ngen=5 x 107), temperature parameter (temp=2) and sampling frequency of 2000, yielding 25000 trees per run. The log files from

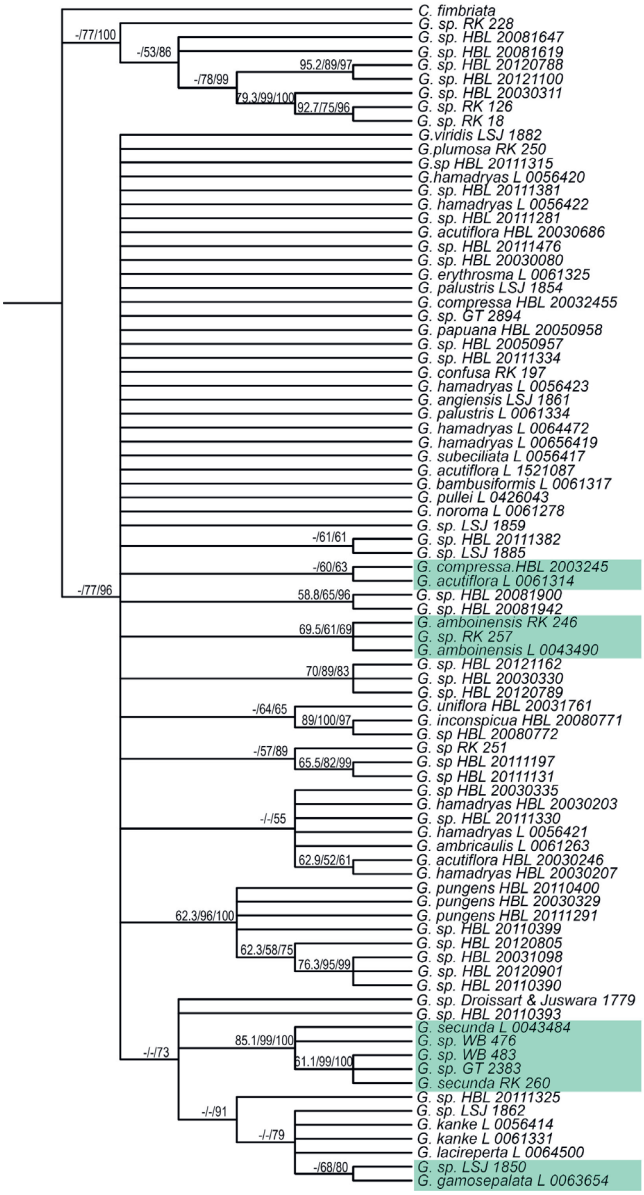

Figure 3.1. Bayesian phylogenetic tree built from aligned sequences from the nrITS region of 81 species and one outgroup. Nodal support values are Neighbour Joining score/bootstrap percentage of Maximum Likelihood/Bayesian posterior probabilities.”-“ indicates bootstrap value <50%. Green colored clades contain herbarium specimens analyzed in this study.

MrBayes were inspected in Tracer v.1.6 to check for convergence of independent runs (i.e., with estimated sample size (ESS)>200).

3.4 Results

3.4.1 Alignment and sequence characteristics

Double-stranded amplifications and complete sequences of nrITS were obtained from 43 fresh specimens of *Glomera*. A total of 38 incomplete sequences were obtained from herbarium specimens. The length of the nrITS sequences varied between 70-887 bp (see Figure 3.2). The oldest specimen investigated was a 195 years old type specimen of *Glomera erythrosma* Blume, from which a 410 bp sequence was recovered, while the youngest was a 4 years old specimen of *Glomera* sp., from which a 404 bp sequence was generated. The length of the DNA sequences obtained from fresh specimens ranged from 706-852 bp (see Figure 3.3).

3.4.2 Phylogenetic analysis

Trees obtained by NJ, ML and Bayesian analyses showed congruent topologies. Figure 3.1 shows the Bayesian tree. Four clades containing sequences obtained from herbarium type specimens were highly supported (>50%) in all three analyses. By matching DNA barcodes, we were able to identify sterile living collections to the following four species of *Glomera*: *G. acutiflora* (Schltr.) J.J.Sm., *G. amboinensis* (Ridl.) J.J.Sm., *G. gamosepalata* P. Royen, and *G. secunda* J.J.Sm. This added

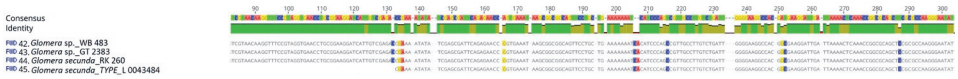

Figure 3.2. Part of the sequence alignment of the nrITS region for different specimens of *G. secunda* J.J.Sm. analysed. The colored nucleotides indicate diagnostic basepairs for this particular species.

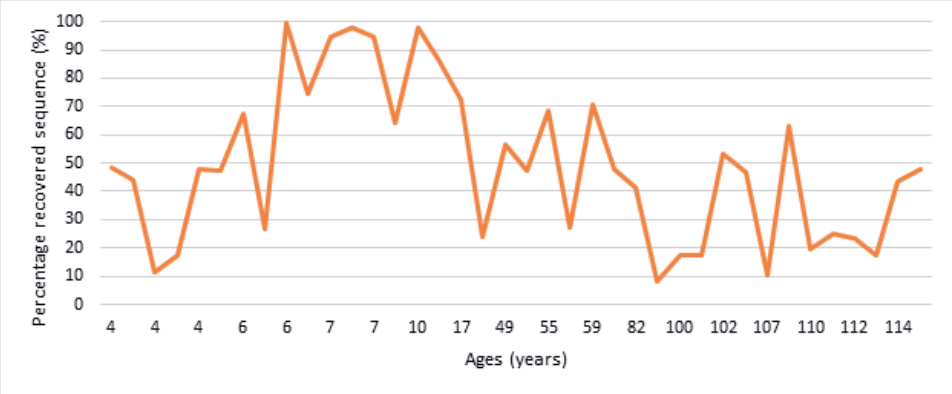

Figure 3.3. Relative recovery of nrITS DNA barcode sequence (%) from herbarium specimens investigated in this study plotted against specimen age (years).

new locality data to their known distributions.

3.5 Discussion

Traditionally, orchid identification heavily relied on morphological characters, both vegetative and floral (Haider et al., 2010). Some of the vegetative (habitus, size, leaf shape) and fertile characters (flower color and size) of *Glomera* species are depicted in Figure 3.4. It is challenging to identify sterile plants to species level as traditional identification keys usually focus on flowering material (Ramalho et al., 2018). DNA barcoding offers a solution to provide rapid and accurate identification using short standardized gene regions (Kress et al., 2005). We show that for the orchid genus *Glomera*, DNA barcoding can be a solution for sterile plant identification. However, not all DNA sequences generated from sterile specimens could be matched with those of type specimens in this study. The still unidentified life plants might belong to undescribed species but could not yet be formally published because flowers are needed for this.

In this study, we show that only small amounts of leaf material (1 cm2) of herbarium specimens were sufficient to generate DNA barcodes, which were sufficiently long and informative to identify sterile specimens in living orchid collections. Especially with species that have many leaves, such as *Glomera*, the sampling of dried specimens can be kept to a minimum. The percentage of recovered DNA sequences was highly variably and no correlation with the age

Figure 3.4. Vegetative and floral characters of *Glomera*. Left: *Glomera inconspicua* J.J.Sm. (HBL 20080772), photograph by Richa Kusuma Wati. Right: *Glomera confusa* J.J.Sm. (HBL 20030288), photograph by Rogier van Vugt.

of the specimen was found. The degradation of DNA in a herbarium is affected by many factors such as the collector's treatment in the field, how rapidly a plant was dried and the storage condition (Hart et al., 2016). A total of 16 of the herbarium specimens analyzed in this study were provided by the Herbarium Bogoriense. DNA sequences recovered were relatively low despite the fact that these specimens were only collected in the past 10 years. The low coverage could be explained by the common method to preserve specimens by compressing the plant and soak it in ethanol 70% in the field before subsequent drying in an oven. The use of ethanol in the tropics, where the high humidity often prevents efficient and continuous dessication, has a negative effect on plant DNA quality (Staats et al., 2011).

The challenges of working with ancient DNA from herbarium material is that only short fragments can be amplified, due to fragmentation caused by ageing. An increasing number of studies show that DNA barcoding is possible from herbarium specimens though (Bakker et al., 2016; Contreras-Ortiz et al., 2019) by designing new primers, as done in this study, or using Next Generation sequencing. The added value of using herbarium collections is that detailed locality data are available (Xu et al., 2015), saving costs for travel to resample. The entire collection of 3095 herbaria in the world today consist of 387 million specimens,

collected from all over the world, and including many rare and possibly already extinct species. Only a small part of these collections are used for DNA-based research, mainly due to the fact that curators are often reluctant to allow destructive sampling (Särkinen et al., 2012; Thiers, 2019). By opening up these collections for DNA barcoding, reference libraries of DNA sequences from especially types could be a welcome solution to help solving the current taxonomic impediment.

3.6 Conclusions

Despite their often degraded DNA, exploitation of DNA barcodes generated from type specimens seems a promising tool to identify sterile specimens in living orchid collections. Additional sequencing of type specimens for DNA barcoding will increase our knowledge of poorly known orchid genera.