EXTRACTION AND AMPLIFICATION OF DNA FROM ORCHID EXSICCATES CONSERVED FOR MORE THAN HALF A CENTURY IN A HERBARIUM IN BOGOTÁ, COLOMBIA

LAURA C. MAZO1, ALBERTO GÓMEZ1,2,4, SONIA R. QUINTANILLA1,2, JAIME E. BERNAL1,2 & PEDRO ORTIZ VALDIVIESO †2,3

1 Instituto de Genética Humana, Facultad de Medicina, Pontificia Universidad Javeriana, Cra 7 # 40-62, edificio 32, Bogotá D.C., 110231, Colombia
2 Iniciativa Genómica Javeriana, Pontificia Universidad Javeriana, Cra 7 # 40-62, edificio 32, Bogotá D. C., 110231, Colombia
3 Asociación Bogotana de Orquideología, Avenida calle 63 # 68G-14, Bogotá D. C., Colombia
4 Corresponding author: agomez@javeriana.edu.co

ABSTRACT. Plant tissue from herbarium specimens contains DNA that has undergone post-mortem degradation. Only small amounts of possibly degraded genetic material free of chemicals and impurities can be extracted from these samples. The aim of the present work was to compare and determine which one of three previously published plant DNA extraction protocols would extract good quality DNA from orchid herbarium specimens stored up to 63 years, susceptible of PCR amplification and Sanger sequencing. The most effective protocol for this type of samples allowed us to obtain and sequence amplified products in 6 of the 7 samples studied with an optimal DNA / protein purity relationship.

RESUMEN. Los tejidos vegetales provenientes de muestras de herbario contienen ADN que ha sufrido degradación post-mortem. Solo pequeñas cantidades de material genético eventualmente degradado, libre de químicos e impurezas, puede ser extraído de dichas muestras. El presente trabajo buscó comparar y determinar cuál de los tres protocolos de extracción seleccionados, que habían sido publicados previamente para ADN vegetal, permitía extraer ADN de buena calidad, susceptible de ser amplificado y secuenciado, a partir de exsicados de orquídeas conservadas en herbario hasta por 63 años. De los protocolos evaluados, el más eficaz permitió obtener y secuenciar productos amplificados en 6 de las 7 muestras estudiadas, con las mejores relaciones de pureza ADN / proteína.

KEY WORDS: DNA, aDNA, Orchids, matK, Exsiccates, Herbarium

Introduction. The term “ancient DNA” (aDNA) refers to the DNA extracted from fossil remains and individuals preserved in musea and herbaria (Andreasen et al. 2009). Plant tissue from herbarium specimens contains DNA that is eventually degraded (Gugerli et al. 2005) with fragments not bigger than 500 base pairs (Audic et al. 2002). Moreover, the genetic material of those specimens can be: a- Contaminated with DNA of microorganisms such as fungi or bacteria (Deagle et al. 2006); b- Accompanied with precipitated chemical material, including polyphenols, terpenoids, tannins and polysaccharides (Huang et al. 2002), which can act as inhibitors of Taq DNA polymerase in plant DNA amplification protocols. A third challenge for working with ancient DNA is a process called cytosine deamination, which creates erroneous DNA sequences (Hofreiter et al. 2001).

Achieving adequate taxon sampling can often be challenging in order to perform molecular phylogenetic studies in plants, due to various reasons such as: a) high costs of field work in remote locations; b) difficulty in obtaining collection permits and access to genetic resources; c) inaccessibility to politically unstable regions; d) difficulty of finding the target species; e) local or global extinction of the species; and f) old name bearing type material lacking diagnostic characters may be included in molecular phylogenetic studies to investigate links with more recently collected specimens (Puillandre et al. 2012).
Collections found at herbaria are a potential source of DNA for improving taxa sampling, and thereby improve current phylogenetic understanding when there is no recent material available (Bebber et al. 2010). In addition, herbarium specimens are considered material with a high historical value due to the antiquity of the samples, and also to their character of taxonomic and biogeographic reference. Small amounts of tissue can be removed from the samples without causing damage to their global integrity. For this reason, the revision of extraction methods to recover as much DNA as possible from the minimum amount of tissue is crucial (Rohland & Hofreiter 2007).

A wide range of techniques have been published to date, seeking to increase the amount of DNA available and reduce the influence of PCR inhibitors. Among these, two main methods have been widely used on both fresh and herbarium samples: a- Cetyl-Trimethyl-Ammonium Bromide or CTAB (Doyle and Doyle 1987) and b- DNeasy Plant Mini Kit from Qiagen (Cota-Sánchez et al. 2006). Silica gels have been included by Van Geel et al. (2008), as a useful supplementary step in the purification of amplified ancient plant DNA. These protocols have been proposed for the extraction of ancient DNA free of chemicals and impurities from herbarium specimens, but have not been widely tested on orchids and are limited to particular groups of plants (Cota-Sánchez et al. 2006).

Along these lines, this study aimed to compare and evaluate three (3) modified extraction protocols, two of which were CTAB based, on the following species Prosthechea grammatoglossa, Epidendrum secundum and Oncidium ornithorrhynchum, collected between 1948 and 2011, to determine the best protocol for DNA extraction and amplification from orchid samples preserved in herbaria, according to previous reports which were applied to newly collected plants.

**Material and methods**

**Selection of plant material.** – DNA was extracted from 7 samples belonging to 3 species of orchids. Of these, 5 were dry tissue (Table 1) and the other 2 corresponded to fresh tissue samples of the species Prosthechea grammatoglossa and Epidendrum secundum. The reason why these specimens were chosen, was to include in our protocols preserved orchid samples with the longest available period of conservation, stored in the Universidad Javeriana Herbarium, to be compared with fresh material of the same species. We worked with 2 mg and 5 mg of tissue corresponding to the homogenization of subsamples of petal, stem and leaves, due to the small amount of material available.

**DNA extraction.** – At first, Petri dishes were washed with distilled water, sodium hypochlorite and absolute ethanol before being used and when switching from one sample to another. Then, 2 mg and 5 mg of tissue were weighed and pulverized by means of a surgical blade. Subsequently three (3) different plant DNA extraction protocols were assessed: Quintanilla’s et al. (2011), Cota-Sánchez’s et al. (2006) (Modified: Ethanol 70% and 99%, no addition of RNAse) and Jobes et al. (1995) (Modified: LiCl was substituted by Ammonium Acetate). The reason why these three protocols were chosen was, first, to reduce the variables to be considered, second, to employ

<table>
<thead>
<tr>
<th>Species</th>
<th>Description</th>
<th>Location</th>
<th>Elevation</th>
<th>Collector</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidendrum secundum Jacq</td>
<td>Different colors according to the land or varieties of the species</td>
<td>Villa blanca (Sasaima, Cundinamarca)</td>
<td>1400 m</td>
<td>C. Ortiz, S.J</td>
<td>17 Oct 1948</td>
</tr>
<tr>
<td>Epidendrum secundum Jacq.</td>
<td>Rupicola. Lilac flowers</td>
<td>Vereda Vancouver (Santanderito, Cundinamarca)</td>
<td>1900 m</td>
<td>P. Ortiz, S.J</td>
<td>3 Febr 1968</td>
</tr>
<tr>
<td>Oncidium ornithorrhynchum H.B.K.</td>
<td>Epiphyte with yellow flowers and brown clear spots</td>
<td>Hacienda la Esmeralda (Madrid, Cundinamarca)</td>
<td>2650 m</td>
<td>S. Restrepo, S.J</td>
<td>30 May 1968</td>
</tr>
<tr>
<td>Oncidium ornithorrhynchum H.B.K.</td>
<td>Yellow flowers with brown spots</td>
<td>“La Rusia” (Duitama, Boyacá)</td>
<td>2870 m</td>
<td>P. Ortiz, S.J</td>
<td>Jun 1975</td>
</tr>
<tr>
<td>Prosthechea grammatoglossa (Rchb.f) W. E. Higgins</td>
<td>Yellow flowers</td>
<td>“La Rambla” (Santanderito, Cundinamarca)</td>
<td>2000 m</td>
<td>H. Silva</td>
<td>13 Jan 1949</td>
</tr>
</tbody>
</table>
different reagents for DNA extraction such as CTAB in conjunction with Beta-Mercaptoethanol (ME), used in Cota-Sánchez et al. (2006) and Quintanilla et al. (2011) with different periods and incubation times, and Sodium Dodecyl-Sulfate (SDS), Polyvinylpyrrolidone (PVP), Dithiothreitol (DTT), Proteinase K (PK) used in Jobes et al. (1995), third, to minimize the cost of DNA extractions, as supplementary purification steps such as silica columns are expensive. In addition, the protocol proposed by Quintanilla et al. (2011) was included as a reference in our work, as it was previously employed with success by our group on the molecular characterization of different orchids. In this later protocol, a purification step based on ammonium acetate (7.5 M) and 2-Propanol substitutes silica columns.

Spectrophotometry, amplification and sequencing – DNA quality was determined by means of a 260/280 - 260/230 purity relationship. Subsequently, a region of 640 bp corresponding to the gene matK was amplified using the following primers:

19Forward: 3’-CGTTCTGACCATATTGCACTATG-5’
556Reverse: 3’-GAAGAAACATCTTTGGATCCA-5’.

This gene was previously used successfully by our group in the characterization of different species of orchids on fresh material (Quintanilla et al. 2011), and matK is reported as one of the official plant BoLD markers as follows: “The BoLD Identification System (IDS) for rbcL and matK is the default identification tool for plant barcodes and accepts sequences from the Ribulose-bisphosphate carboxylase and Maturase K genes” (BoLDsystems, 2012), and also as the reference for the CBOL Plant Working Group (Janzen 2009).

PCR reactions contained 6µl of PCR buffer (5x), 3 µl of each primer (10mM), 0.9 µl of MgCl (2.5 mM), 0.6 µl of dNTPs (200 µM), 0.3 µl of DNA Taq polymerase (1.25 units) and 14.4 µl of sterile water to complete a volume of 30 µl of each reaction. PCR cycles were programmed as follows: 3 min at 95°C initial denaturation, 39 cycles of 30 s at 95°C, 30 s at 50.4°C, 48 s at 72°C and a final extension of 5 min at 72°C. After amplification, DNA concentration was determined on 7μl of PCR product through electrophoresis in 1% agarose and 1x TBE buffer. The amplified products were Sanger sequenced (Applied Biosystems 3730XL at Macrogen, Korea) and analyzed using the programs Chromas Lite and BioEdit with 5 other sequences reported in GenBank (Table 2) and finally a phylogenetic tree was drawn through the Neighbor-Joining method.

Statistical analysis. – A descriptive statistical report was produced by means of categorizing the results from the different indices. This was performed dividing DNA concentration values as: a) less than 10 ng (indicating low extraction, inefficient protocol; insufficient tissue quantity); b) between 11 and 30 ng (corresponding to medium extraction and an average efficient protocol) and, c) ranging from 31 to 86 ng (corresponding to an optimum extraction, a very efficient protocol and a sufficient quantity of tissue), while 260/280 and 260/230 absorbance relations were classified as: a) less than 1.8 (indicating a bad quality, and contamination by proteins); b) between 1.8 and 2.2 (good quality, free of contamination) and, c) over 2.2 (bad quality, contaminated by ARN, polyphenols and polysaccharides) (Cattaneo et al. 2006).

Results and discussion
Due to the limited amount of nucleic acids preserved in herbarium specimens, it is difficult to obtain an appropriate amount of molecules in the early stages of extraction. For this reason, an appropriate lysis buffer is a key factor for an optimal obtention of DNA molecules (Yang et al. 1997). The use of the Cetyl-Trimethyl-Ammonium Bromide (CTAB) extraction protocol was designed as an efficient method to avoid contamination with polysaccharides (Thine & Telle 2008). This method was evaluated in the protocols proposed by Quintanilla et al. (2011) and Cota-Sánchez et al. (2006), both of which include the use of reagent Beta-Mercaptoethanol (ME), which is a reducing agent that blocks the action of phenols and, when used at higher concentrations (1%), seems to be more efficient in reducing oxidation (Mittmann et al. 2007). Although in both Quintanilla et al. (2010) and Cota-Sánchez et al. (2006) an equivalent concentration of

<table>
<thead>
<tr>
<th>Table 2. Sequences reported in GenBank.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oncidium ornithorhynchum</strong></td>
</tr>
<tr>
<td><strong>Oncidium unguiculatum</strong></td>
</tr>
<tr>
<td><strong>Oncidium incurvum</strong></td>
</tr>
<tr>
<td><strong>Peristeria elata</strong></td>
</tr>
</tbody>
</table>

ME was used, there is a pronounced difference in their efficiency, as the protocol proposed by Quintanilla et al. (2011) results in higher quality DNA (Table 3).

In contrast, the buffer evaluated in the protocol proposed by Jobes et al. (1995) is composed of Sodium Dodecyl-Sulfate (SDS), Polyvinylpyrrolidone (PVP), Dithiothreitol (DTT), Proteinase K (PK) and no ME. The PVP and DTT are added in order to absorb polyphenols, which tend to co-precipitate along with DNA (Mittmann et al. 2007). In spite of this, it is evident that the use of ME in the lysis buffer allowed to achieve better results than the use of PVP and DTT, because there was no amplification with any amount of tissue (2 mg or 5 mg) using the protocol of Jobes et al. (1995) in any of the seven (7) species studied (Fig. 1-2).

Another reagent used in this buffer is proteinase K (PK), which’s function is to digest proteins, and it is important to remove the Ethylene-Diamine-Tetra-Acetic Acid (EDTA) from the initial incubation steps, because this reagent inhibits the function of the PK (Cattaneo et al. 2006). Also, it is necessary to denature the proteinase K afterwards as this reagent may in turn inhibit the PCR (Orouke et al. 1996). The protocol proposed by Jobes et al. (1995) does not indicate any steps related to the elimination of EDTA before using the PK neither its subsequent inactivation prior to PCR. This may interfere with the action of PK during the initial incubation and that is probably the reason why there was no amplification of any of the species studied (Cattaneo et al. 2006).

During the DNA extraction it is important to consider steps for incubation and precipitation, as it seems unnecessary and specially inefficient to extend and increase the times and temperatures in order to obtain more DNA (Mittmann et al. 2007). Sablok et al. (2009) observed that incubation periods with temperatures above 65°C (without reaching 80 °C) resulted in loss of DNA. In the protocols proposed by Jobes et al. (1995) and Cota-Sánchez et al. (2006), the incubation temperature varied between 50°C - 56°C and in Quintanilla et al. (2011) between 60°C - 65°C. This contrasts with the observations made by Sablok et al. (2009) because at high temperatures more DNA was obtained, being evident to achieve higher quantities of DNA concentrations upon application of the Quintanilla’s et al. protocol (Table 3).

On the other hand, incubation of nucleic acid / salt / ethanol performed at low temperatures (i.e. -20°C) is quoted in several articles. Nevertheless, this is not necessary since nucleic acids precipitate at a temperature of -4°C in concentrations below 20 ng/ml after 15-30 minutes (Bitesize Bio). In the protocol of Jobes et al. (1995), four incubations were performed, three of them at -20°C for 30 minutes and the last one at

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosthechea grammatoglossa (1949)</td>
<td>2</td>
<td>32,5</td>
<td>2,04</td>
</tr>
<tr>
<td>Oncidium ornithorrhynchum (1975)</td>
<td>2</td>
<td>28,7</td>
<td>1,98</td>
</tr>
<tr>
<td>Oncidium ornithorrhynchum (1968)</td>
<td>2</td>
<td>5,3</td>
<td>2,42</td>
</tr>
<tr>
<td>Epidendrum secundum (1968)</td>
<td>2</td>
<td>15,1</td>
<td>1,99</td>
</tr>
<tr>
<td>Epidendrum secundum (1948)</td>
<td>2</td>
<td>6,7</td>
<td>2,67</td>
</tr>
<tr>
<td>Prosthechea grammatoglossa (2011)</td>
<td>2</td>
<td>4,6</td>
<td>3,54</td>
</tr>
<tr>
<td>Epidendrum secundum (2011)</td>
<td>2</td>
<td>7,4</td>
<td>2,22</td>
</tr>
<tr>
<td>Prosthechea grammatoglossa (1949)</td>
<td>5</td>
<td>75,8</td>
<td>1,52</td>
</tr>
<tr>
<td>Oncidium ornithorrhynchum (1975)</td>
<td>5</td>
<td>85,6</td>
<td>1,66</td>
</tr>
<tr>
<td>Oncidium ornithorrhynchum (1968)</td>
<td>5</td>
<td>-2,9</td>
<td>1,51</td>
</tr>
<tr>
<td>Epidendrum secundum (1968)</td>
<td>5</td>
<td>57,2</td>
<td>1,91</td>
</tr>
<tr>
<td>Epidendrum secundum (1948)</td>
<td>5</td>
<td>28</td>
<td>2,02</td>
</tr>
<tr>
<td>Prosthechea grammatoglossa (2011)</td>
<td>5</td>
<td>53,3</td>
<td>2,03</td>
</tr>
<tr>
<td>Epidendrum secundum (2011)</td>
<td>5</td>
<td>57</td>
<td>1,99</td>
</tr>
</tbody>
</table>
-20°C for one night, unlike Cota-Sánchez et al. (2006) where two incubations were performed at -20°C for two nights, and Quintanilla et al. (2011) where there was only one precipitation at -20°C for one night. This indicates that the smaller number of incubations and precipitations, involving a smaller number of washes, allow the obtention of higher concentration of DNA.

Another issue in the DNA extraction is the removal of polysaccharides, although high concentrations of Sodium Chloride (NaCl) appear to block the action of these polysaccharides (Mittmann et al. 2007). The highest concentrations of NaCl in the lysis buffer were reported by Quintanilla et al. (2011) and Cota-Sánchez et al. (2006), the first being the one that gave better results, unlike Jobes’ et al. protocol (1995) which did not yield any amplified product from any species (Fig. 1-2), and whose total NaCl concentration was 500 mM. Chang et al. (1993) state that increasing salt concentration from 1.4 M to 2.0 M in the lysis buffer, affects the solubility of the polysaccharides in ethanol, allowing optimum precipitation of nucleic acids (Mittmann et al. 2007) which improves the relation 260/230 (Josquin et al. 2006) (Table 3).

Salts such as Sodium Acetate and Ammonium Acetate, are also employed to precipitate DNA. When in solution, Sodium Acetate becomes Na⁺ and [CH₃COO]⁻, and Sodium ions positively charged neutralize the negative charge of PO₃⁻ of nucleic acids, then DNA precipitates more easily (BietsizeBio). In the protocols of Jobes et al. (1995) and Cota-Sánchez et al. (2006), Sodium Acetate was used without obtaining good results. In Quintanilla et al. (2011) Ammonium Acetate was used, and showed a better performance (Fig. 1–2).

Also, several ethanol washing steps are crucial as salt residues are removed from DNA (Bitsize Bio). Regarding Cota-Sánchez et al. (1995) and Jobes et al. (1995) only one wash was performed, contrary to Quintanilla et al. (2011) where there were not only two washes with 70% ethanol, but also two washes with 99% ethanol, indicating that more than one wash with an alcohol leads to purest DNA, as seen in the 260 / 280 relations that are notoriously better in this last protocol than in the previous ones.

It is important to emphasize that the process of degradation and deterioration is unique in each exsiccate, and that particular chemical compounds of...
each herbarium sample can either interfere or favor certain steps during the extraction process (Yang et al. 1997). That is why, in samples belonging to the same species, quite different DNA concentrations can be obtained (Table 1). This may be due to different methods (chemical - use of ethanol or formaldehyde at 30%-, and physical - drying at 70°C-) that each sample endures during its treatment after it is collected (Savolainen et al. 1995). The fact that these species might not be comparable in terms of the presence of particular PCR inhibiting secondary metabolites, could at least partly explain the results obtained on different specimens in our report.

Usually, the length of the sequences obtained from ancient DNA are smaller than 200 bp; in this study we observed that the sequences corresponding to the species collected in 2011, that is: Epidendrum secundum extracted with the protocol Quintanilla et al. (2011) and E. secundum (C+, from 450 mg of tissue), were the most readable sequences with defined peaks (Fig 3–4). The same species, when preserved for a longer time (1948) and compared with the one collected in 2011 and extracted using the protocol Cota-Sánchez et al. (2006), showed sequences which were not easily readable and therefore showed poor quality (Fig. 5–6). The sequences corresponding to the species Oncidium ornithorrhynchum, despite being poorly readable, were similar to each other (Fig. 7–8) and also coincide with the sequence previously reported by Quintanilla et al. (2011) which was obtained from a fresh sample (Fig. 9). Andreasen et al. (2009) propose that the length and quality of the amplified fragment is related to the degree of DNA degradation: the greater time of collection, the greater the DNA fragmentation. This is not observed in the species E. secundum collected in 1968, which although showing a high concentration of DNA, failed to be amplified in contrast to the same species collected in 1948, which was easily amplified (Fig. 1–2). This result can be influenced by the treatment that was given to each herbarium specimen when collected and preserved. However, in both cases, their corresponding data were not noted and kept, so there is no available information that would allow to perform a correlation. Hanni et al. (1994), suggest that inefficient amplification of the samples is due to the small amount of tissue used corresponding to 2 mg and 5 mg as, in other cases, approximately 30 mg or even 100 mg of plant tissue have been used (Lister et al. 2008).
In this study, DNA extracted using the protocol Quintanilla et al. (2011) and 5 mg of tissue, showed a higher concentration of Nucleic Acids than those obtained from 2 mg (Table 1). Therefore, the higher amount of available material, the greater the possibility of obtaining higher concentrations of DNA.

Finally, the Neighbor-Joining method (NJ) is frequently used in studies of “molecular barcode” (DNA barcoding) when aiming to identify unknown individuals and to generate a single phylogenetic tree with the lowest genetic distances (Saitou & Nei 1987). Thus, by reflecting the similarity between sequences, individuals belonging to the same species will integrate isolated groups due to their similarities (similarity in terms of sequences) (Peña 2011). We can then infer that the sequences of the species Oncidium ornithorrhynchum form an isolated group that also includes the previously reported sequences of this same species in NCBI (Fig. 11).

In conclusion, the methodology for DNA extraction on herbarium individuals with the protocol proposed by Quintanilla et al. (2011) is effective for samples with storage times over fifty years, as compared with the other two methodologies reported and studied in the present work on this type of specimens. The importance of prior information available from the herbarium related with the method of fixation (chemical or physical) of each individual has to be emphasized to assess how the samples will eventually work, as it was found upon completion of this study that the effect of time is not as important for DNA degradation as the initial use of formaldehyde, methanol or application of heat. We recommend: 1- Further testing different sets of primers that may generate shorter fragments, 2- DNA polymerases that have the property to correct readings, and 3- Next generation sequencing to obtain more replicates of ancient DNA sequences to reveal deamination generated errors (Metzker 2010).

ACKNOWLEDGMENTS. The authors thank the staff from the “Lorenzo Uribe, SJ” herbarium (Faculty of Science, Pontificia Universidad Javeriana), as well as the staff of the Institute of Human Genetics (Faculty of Medicine, Pontificia Universidad Javeriana), particularly our colleagues involved in the research group “Ancient DNA”. We also thank CorpoGen for the use of their DNA quantification equipment. Our final manuscript includes important remarks from two anonymous reviewers, to whom we are indebted.

Figure 9. Part of the sequence of the gene matK of the species Oncidium ornithorrhynchum, reported in 2010 by Quintanilla et al. (H02192561).

Figure 10. Part of the sequence of the gene matK of the species Prosteecha grammatoglossa 2011 (Quintanilla et al. 2010).

Figure 11. Genetic distance phylogram through Neighbor-Joining clustering of the sequences studied plus 6 previously reported sequences in GenBank.
Literature cited


