# SYNTHESYS 3

## Synthesis of Systematic Resources

A Seventh Framework Programme funded project

**Project Number** 312253

### DELIVERABLE 2.8

# Protocols for DNA extraction: Protocols for DNA extraction made available

**Date:** December 13, 2016

**Work package number** WP2. Improving collections management and enhancing accessibility.

**Objectives 2:** Developing strategic priorities for molecular related NH collections

**Task 2.3**. Develop protocols for data collection from DNA extraction

"Develop protocols for collecting data from sequencing activities on NH collections and feedback on the success of different methodologies. Working alongside the CPB, NA2 aims to produce a Memorandum of Understanding for a European protocol"

**Deputy Partner:** CSIC

**Deliverable author(s):** Isabel Rey Fraile, Beatriz Álvarez Dorda, Inés Álvarez Fernández, Javier Fuertes Aguilar, Elisabeth Haring, Virginie Bouetel, Tim K. Fulcher, Thomas von Rintelen, René Dekker, Jacqueline Mackenzie-Dodds, Michelle Hollingsworth, Frantisek Vacek, Anna Hundsdörfer, Nathalie Smitz, Larissa Smirnova, Virgilio Massimiliano.

**CONTENTS**

## 1. INTRODUCTION

The SYNTHESYS Networking Activities (NAs) are an integrated package that support and develop the natural history infrastructure in Europe. The more integrated the actions resulting from these activities, the better the experience of SYNTHESYS Access users, the exchange between institution, and the general public access will be.

The NAs aim to raise awareness of good practices in handling and sampling collections by offering guidelines for the care, storage and conservation of collections, ensuring usability of collections. This practice is reinforced by the increased adoption of common standards and protocols. As a result, an integrated European resource will arise bringing together the biological and geological collections held by major Natural History (NH) museums and other institutions.

This report refers to the sub-task 2.3 ("Develop protocols for collecting data from sequencing activities on NH collections and feedback on the success of different methodologies") within objective 2 named "Developing Strategic Priorities for molecular NH collections" linked to the WP2 "Improving collections management and Enhancing accessibility".

All Natural History collections in SYNTHESYS project pursue the same common aims: preservation, uses and accessibility. Thereby, the main objective of Tissues and DNA Collections, also known as biobanks, in these institutions are to preserve and make accessible the molecular genomic biodiversity (both physical samples/specimens, either extinct or extant and information data) to the maximum number of beneficiaries on short, medium and long-term.
In the 1980s, nucleic acid-based research reached all disciplines, including the non-molecular such as taxonomy and phylogeny, housed in museums and other institutions with traditional NH collections. As a result, a new type of collections was created and the possibility of giving new uses to classic collections was opened (Dessauer & Hafner, 1984; Dessauer et al., 1990; Sherwin, 1991; Thomas, 1994, Rey, 2013, 2014). These new collections specifically designed as source of DNA, which mainly kept samples dry or frozen, were consolidated as an important scientific infrastructure for genomic and genetic studies in the current century (Prendini et al., 2002; Savolainen et al., 2006). Furthermore, these collections became vital to preserve orphaned molecular samples remaining after the finalization of many research projects. From a quantitative perspective, the number of samples (tissues and DNA) managed by specific laboratories or research teams should normally be less than in biobanks. However, as common also to most SYNTHESYS institutions, in reality the set of samples from the different projects is usually much larger than the samples kept in the biobanks and the lack of professional management of these will cause problems in future and. Here, we consider a biobank as defined by ISBER -International Society for Biological and Environmental Repositories- (Hewitt & Watson, 2013). In order to be considered as such, the data and samples must be associated with a database backup with defined standards and the curatorial work within the collection must be carried on by a specialist and professional staff.

From our perspective the protocols, processes and goals from the two types of facilities are different. If we take the definition given above into account, in a molecular laboratory, samples are processed only to obtain specifics results from partial or complete genomes that are normally linked to research goals usually to produce a scientific paper or report. The external use of those samples is not always granted once the project ends. This is common practice also in SYNTHESYS institutions (Biobanks Collection managers and curators, pers. comm.). Due to different temporal frames (long-term vs. short-term), the databases associated with each of these facilities have very different contents. In the case of long-term preservation of collections the main aim in databasing is to retain the information related to preservations, extraction methods and access for futures uses, while the techniques applied (e.g., amplification or genotyping) might not be as relevant (although they could be considered as supplemental data). In contrast, in molecular laboratories, due to their objective-driven nature - e.g., species or gender identification -, the preservation of the molecular extracts may not be a priority and related information (e.g. localities, geo-referencing, preservation history, etc.) may

even not be included in the databases, while this information is of vital importance for NH collections. Experiences have shown that field sampling, the type of death induction (particularly interesting for insects), historical information of specimens or samples (old or modern), preservation and molecular extraction methods, as well as environmental parameters, are crucial for obtaining good quality DNA.

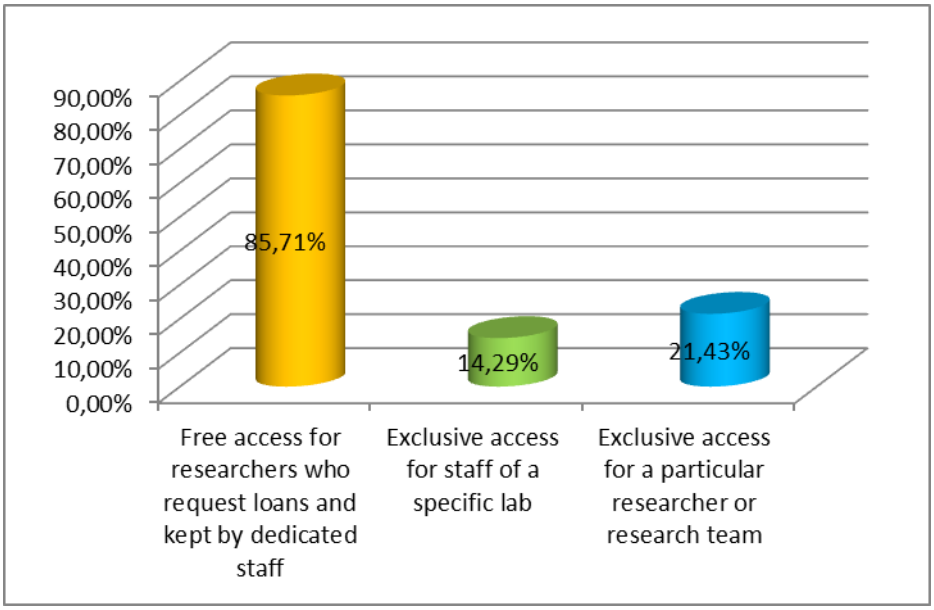
To accomplish this task we aimed to quantify the tissues and molecular collections as well as NH biobanks that exist in the different SYNTHESYS institutions involved, and to investigate what kind of information associated with them is retained. For this reason, a simple survey to determine the state of the art was designed and conducted. DNA collections and preservation of data from DNA extraction were assessed through the answers of an online survey, and by contacting the person in charge of this task at each institution, working in NH molecular collections. The "Survey Synthesys 3 WP2 Task 2.3" collected information on conservation and care of NH molecular collections, available in SYNTHESYS institutions.
The survey was made available online via Survey Monkey (https://www.surveymonkey.com/s/DQLLJNG) from 5 February 2015 to 4 March 2015.
The detailed questionnaire is presented in the Annex I and partners invited to participate are listed in the Annex II.

## 2. DNA COLLECTIONS AND PRESERVATION OF DATA FROM DNA EXTRACTION: WHAT IS THE SITUATION IN SYNTHESYS EUROPE?

### 2. 1. SYNTHESYS INSTITUTIONS WITH TISSUES AND DNA COLLECTIONS AND DATA USERS

**Does your institution have Tissues and DNA Collections?**

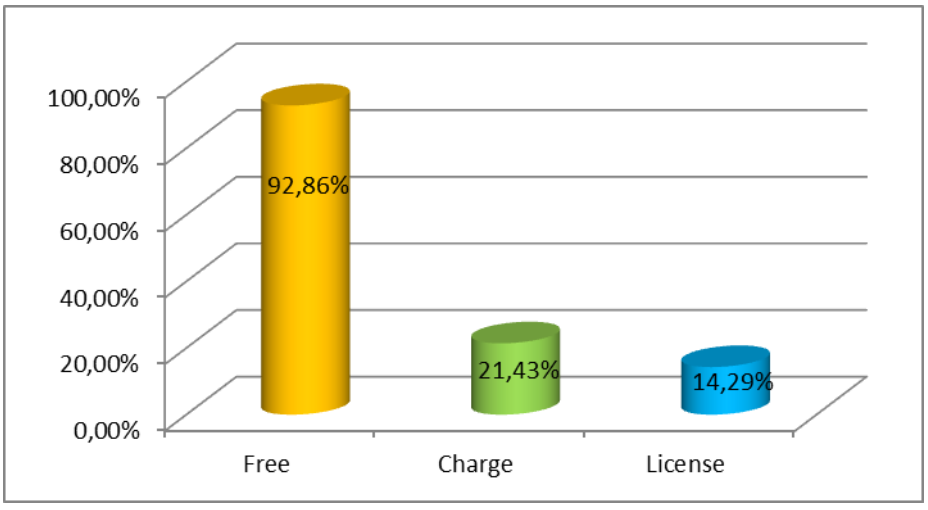
- Answered: 14, Skipped: 2




| Answer Options | Replies | |
|---|---|---|
| **Free access for researchers who request loans and kept by dedicated staff** | 12 | 85,71% |
| **Exclusive access for staff of a specific lab** | 2 | 14,29% |
| **Exclusive access for a particular researcher or research team** | 3 | 21,43% |
| Total number of survey respondents: **14** | | |

**Conditions of use**

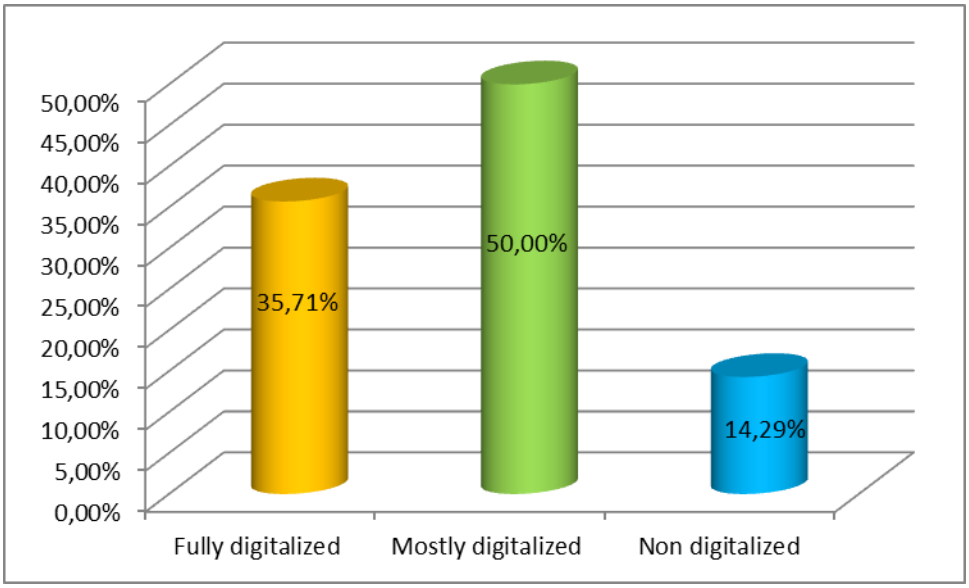
- Answered: 14, Skipped: 2




| Answer Options | Replies | |
|---|---|---|
| **Free** | 13 | 92,86% |
| **Charge** | 3 | 21,43% |
| **License** | 2 | 14,29% |
| Total number of survey respondents: **14** | | |

Is the collection managed through a data base?

- Answered: 14, Skipped: 2

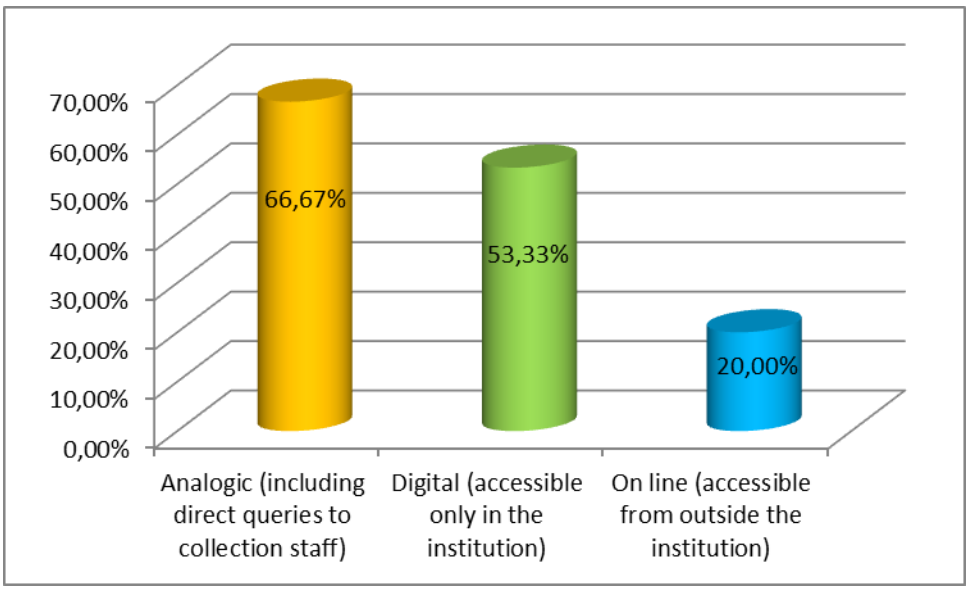



| Answer Options | Replies | |
|---|---|---|
| Fully digitalized | 5 | 35,71% |
| Mostly digitalized | 7 | 50,00% |
| Non digitalized | 2 | 14,29% |
| Total number of survey respondents: 14 | | |

## 2. 2.- INFORMATION INCLUDE IN THE CURRENT DATABASES

Regarding external Access to database

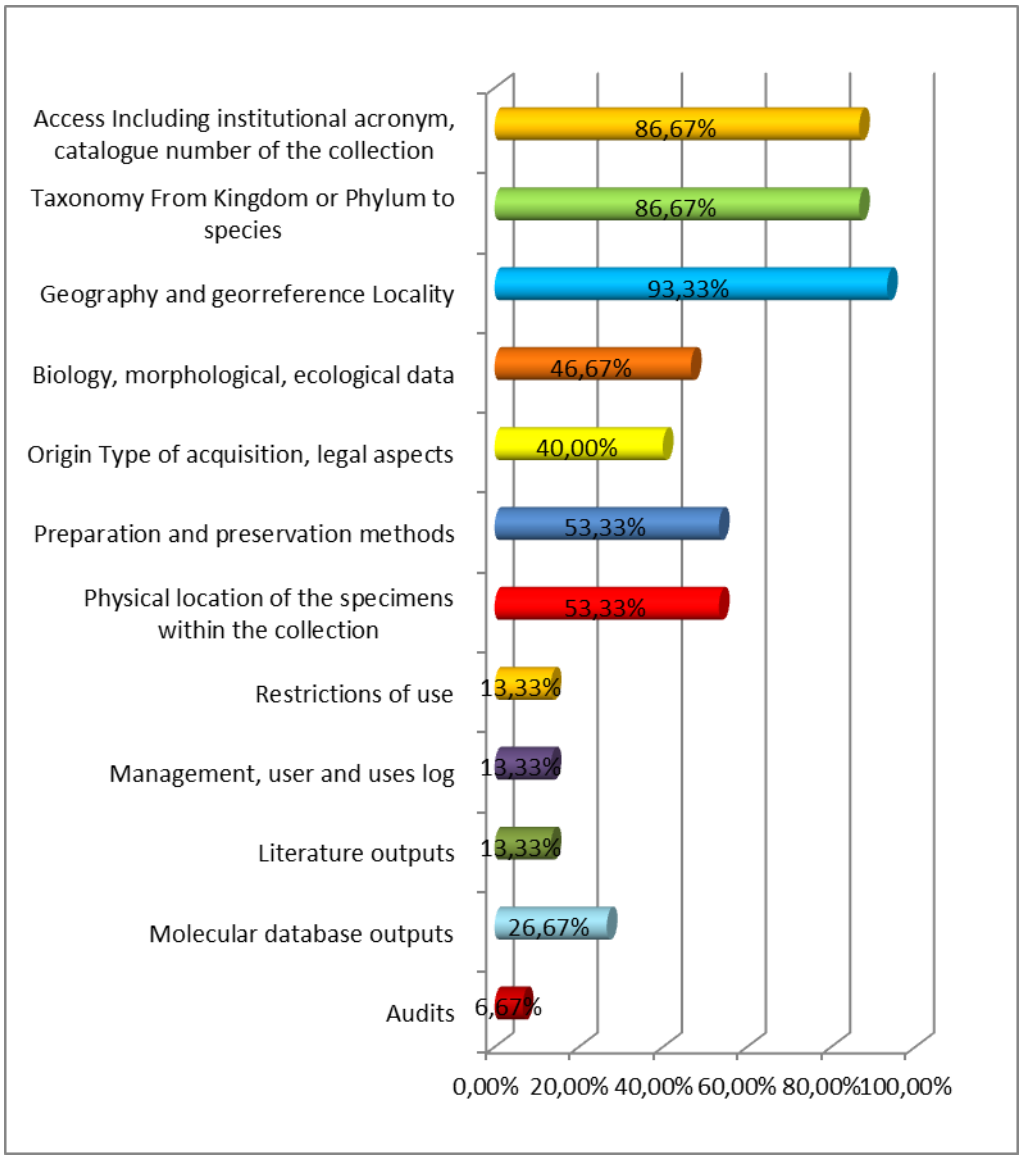
- Answered: 15, Skipped: 1




| Answer Options | Replies | |
|---|---|---|
| Analogic (including direct queries to collection staff) | 10 | 66,67% |
| Digital (accessible only in the institution) | 8 | 53,33% |
| On line (accessible from outside the institution) | 3 | 20,00% |
| Total number of survey respondents: 15 | | |

5. The database includes information on: (choose as needed)

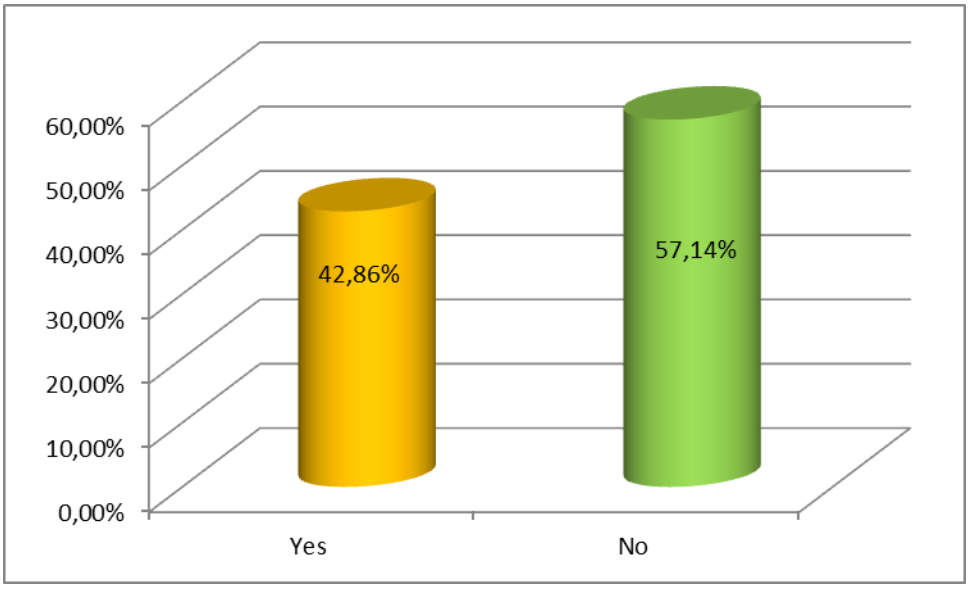
- Answered: 15, Skipped: 1




| Answer Options | Replies | |
|---|---|---|
| Access Including institutional acronym, catalogue number of the collection | 13 | 86,67% |
| Taxonomy from Kingdom or Phylum to species | 13 | 86,67% |
| Geography and georeference Locality where it was collected as exact as possible, including land or marine GPS coordinates | 14 | 93,33% |
| Biology, morphological, ecological data | 7 | 46,67% |
| Origin Type of acquisition, legal aspects ( e.g. collecting, importing or donation permits ) | 6 | 40,00% |
| Preparation and preservation methods | 8 | 53,33% |
| Locations Information on the physical location of the specimens within the collection | 8 | 53,33% |
| Restrictions of use | 2 | 13,33% |
| Management, user and uses log | 2 | 13,33% |
| Literature outputs | 2 | 13,33% |
| Molecular database outputs | 4 | 26,67% |
| Audits | 1 | 6,67% |
| Total number of survey respondents: 15 | | |

**Regarding acid nucleic samples, your collection retains information about the extractions, purifications and preservation methods used?**

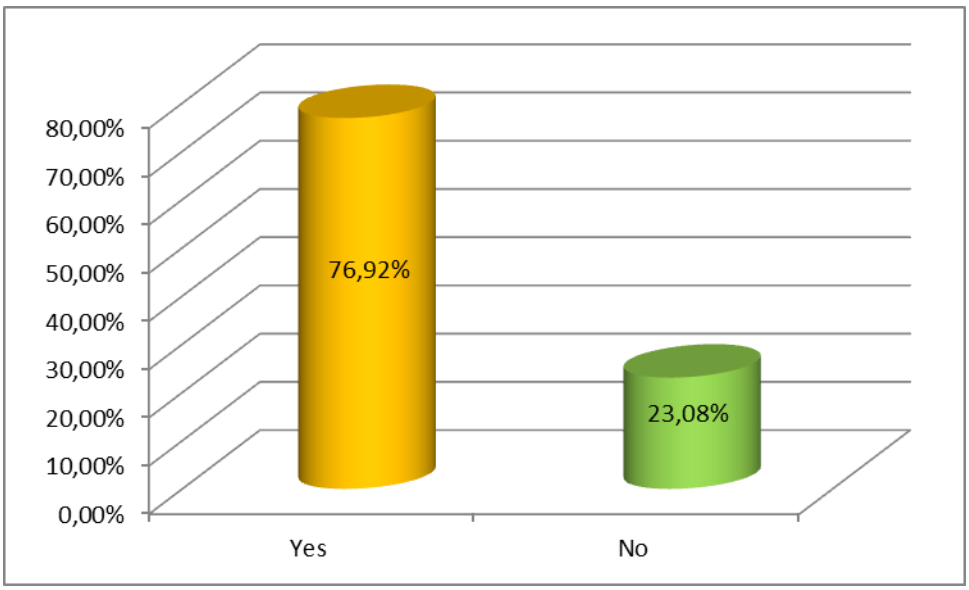
- Answered: 14, Skipped: 2




| Answer Options | Replies | |
|---|---|---|
| **Yes** | 6 | 42,86% |
| **No** | 8 | 57,14% |
| **Total number of survey respondents: 14** | | |

**Regarding tissue samples, your collection retains information about the preparation and preservation methods used?**

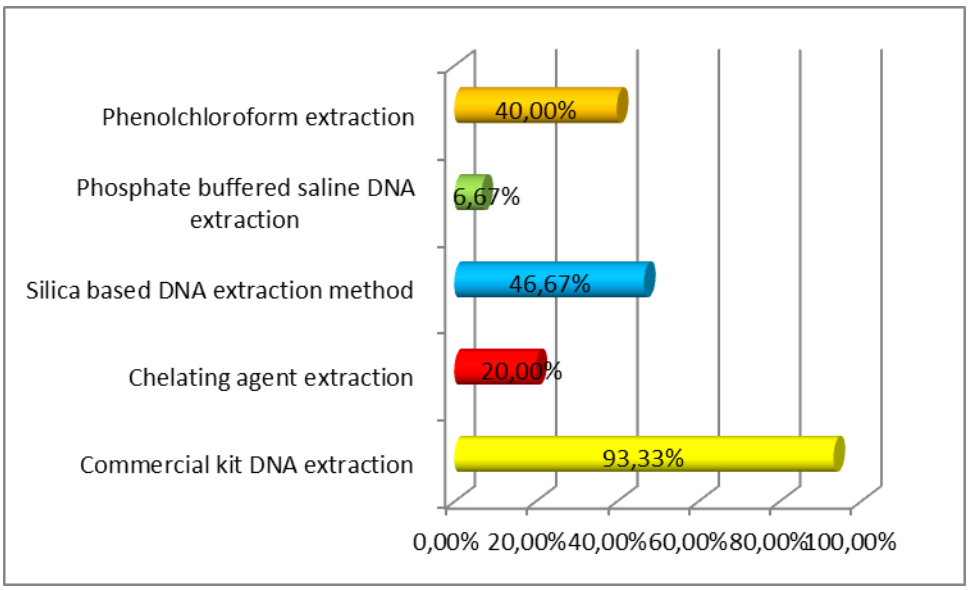
- Answered: 13, Skipped: 3




| Answer Options | Replies | |
|---|---|---|
| **Yes** | 10 | 76,92% |
| **No** | 3 | 23,08% |
| **Total number of survey respondents: 13** | | |

8. Can you select any methods used in the following options?

- Answered: 15, Skipped: 1




| Answer Options | Replies | |
|---|---|---|
| **Phenolchloroform extraction** | 6 | 40,00% |
| **Phosphate buffered saline DNA extraction** | 1 | 6,67% |
| **Silica based DNA extraction method** | 7 | 46,67% |
| **Chelating agent extraction** | 3 | 20,00% |
| **Commercial kit DNA extraction** | 14 | 93,33% |
| **Total number of survey respondents: 15** | | |
| **Comments: 5** | | |

### 3. PROTOCOLS FOR COLLECTING DATA

On the basis of the gathered information from the surveys and the work developed in this activity we present a general guide compatible with the present organization and work flow across SYNTHESYS members. The databases compiling the information from molecular collection (nucleic acids or other biomolecules) should be somewhat similar to those for the classical collections, but also include other specific characteristics.
Currently, the non-profit scientific and educational international association named Biodiversity Information Standards exists (TDWG http://www.tdwg.org/about-tdwg/), and its priority is "the development of standards for the exchange of biological/biodiversity data". The most widely deployed formats for biodiversity occurrence data are Darwin Core (Wieczorek et al. 2012) and ABCD (Access to Biological Collections Data task group. 2010). There is a lot of work of many people and many years behind the development of both standards. Furthermore, the latest standard produced specifically to collect information on genomic biodiversity was done in the framework of GGBN funded by SYNTHESYS III, and can be found in Droege (2016).

The world of molecular samples preservation is very wide and includes numerous stakeholders (e.g., ISBER http://www.isber.org/, EESB -European and Middle Eastern Society for Biopreservation and Biobanking- https://esbb.org/, etc.), and thus several specific databases have been designed, both commercially (Freezerworks http://www.freezerworks.com/) and free of use (VoSeq) (Peña & Malm, 2012). There are also online platforms aiming at facilitating and guaranteeing the management and access to the information, such as Global Biodiversity Information Facility (GBIF, http://www.gbif.org/), Global Genome Biodiversity Network (GGBN http://www.ggbn.org/ggbn_portal/), Barcode of Life Data Systems (BOLD http://www.barcodeoflife.org/), GenBank (https://www.ncbi.nlm.nih.gov/genbank/), the European

Bioinformatics Institute (EBI http://www.ebi.ac.uk/), and the DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp/).

Based on previous work (Rey, 2013, 2014) and the survey on databases for tissue and DNA collections or biobanks, we propose a series of common fields, which can be grouped under the following big areas:

1 ***Access information***
2 ***Taxonomic information***
3 ***Geographical information***
4 ***Biological information***
5 ***Information on the legal origin***
6 ***Information on the preparation and preservation methods***
7 ***Location***
8 ***Restrictions of use***
9 ***Management, record of users and uses***
10 ***Bibliographic records***
11 ***Molecular data outputs***
12 ***Audits***

We assume this is a *desideratum* as in many occasions either due to the lack of information or of resources compiling this information is hard or sometimes impossible. However, the general framework presented here would enable compatibility despite some of the data are missing.

**1 Access information:** Universally unique identifier (UUID) or catalogue number (including acronym of the institution, plus the division, plus unique number).
Collector's number or similar, that might preserve information on other related material or of the specimen (e.g., calls and songs, pictures, field notes, anatomical or histological preparations, or their images).

**2 Taxonomic information:** From phylum to species, including as many taxonomic categories (tribe, subspecies…) as possible. This includes information on whether it is a name-bearing specimen (type specimen) or not aas well as any known synonyms. It is necessary to report the taxonomic references used. For example, the tissue and DNA collection of the MNCN (REY & DORDA, 2006) generally applies:
- For invertebrates the websites http://faunaeur.org/ and http://www.fauna-iberica.mncn.csic.es;
- For classifying vertebrates it follows http://fishbase.org for fishes; Pleguezuelos et al. (2002) and Howard & Moore (1994) for reptiles and amphibians; Madroño et al. (2004) for birds; and Wilson & Reeder (1993) for mammals.

**3 Geographical information:** Collection site as exact as possible, including GPS coordinates on sea or land.

**4 Biological information:** Data collected by researchers when sampling in the field such as the morphological, internal or external characteristics or measurements, sex, age, number of embryos (in case of pregnant females) collected during the necropsy or preparation of the specimen. It also includes information on any existing preparations and related information (e.g., anatomical / histological preparations and its images) and their location.

**5 Information on the legal origin:** This refers to the data of accessioning and provenance, the kind of entry (donation, institutional sampling, seized deposit…), collecting and importation permits, the acceptance reports and any sponsorship documentation.

**6 Information on the preparation and preservation methods**: This should include all the data about the kind of sampling technique, transportation, provisional preservation, type of tissue and type of nucleic acid

obtained. It should reflect the preservation history of tissues and assure a better preservation of both tissues and nucleic acids; and real data when the samples will be requested by users.

### *6.1 Tissues/Plant Fragment*

Type of tissue: skin, feather, liver…
Date of sampling: e.g., DD/MM/YYYY
Original preservation (before sampling): live, refrigerated at 5ºC, frozen at -20ºC or -80ºC, nitrogen, dead. If so, approximate time of death, or in the case of samples from invertebrates, the approximate time of separation (which is relevant in case there is decomposition or fungi growths).
Kind of death: traumatic, stress, asphyxia (water, chemical compound)
Post-sampling preservation**:** refrigerated at 5ºC (e.g., blood, or within an extraction buffer prior to extraction), frozen at -20ºC or -80ºC, nitrogen, buffers (DMSO, other) or solutions (e.g. alcohol, including percentage in composition), dry, freeze-dried (time it remained frozen prior to becoming freeze-dried).
Workflow: defrosting – preservation, including the time of defrosting prior to preserving (can be very important depending on the type of tissue or organism).
Historical collections with classical preservation: dry or spirit.
Time: Date since the preparation has been preserved. Any other previous preservation methods that are known (e.g., remaining in alcohol for 5 years prior to having the dry skin mounted).
In the case of herbaria samples, hours on freezing and/or drying pressing and environmental conditions (humidity, temperature).
For zoological samples, the preservation liquids, chemical products used in the preparation of skeletons or the treatment applied to the skins of the taxidermy specimens.

### *6.2 DNA (sensu lato, as an alternative to nucleic acids or genomes)*

Type: DNA, RNA, genomic, mitochondrial, chloroplast, others.
Extraction method: e.g., extraction kit name, extraction method reference and/or modification about this.
Extraction date: e.g., DD/MM/YYYY.
Purification method: e.g., Purification kit name.
Purification date: e.g., DD/MM/YYYY.
Dilution method: TE, water, buffers.
Volume: e.g., expressed in µl.
Concentration: e.g., expressed in ng/µl and specify the device used to do the measure (e.g., Nanodrop, Qubit, Bioanalyzer etc.).
Working preservation method: refrigerated 5ºC, frozen at -20ºC or -80ºC, room temperature.
Long-term preservation method: frozen -20ºC, -80ºC, nitrogen, dry, freeze-dried (including the period of time frozen prior to being freeze-dried, days or months).
Initial volume for long-term preservation: e.g., expressed in µl or before dried.
Concentration for long-term preservation (if applied).
Containers name: e.g., 2D smartScan barcode boxes.
Company manufacturing containers:
Type of container material: e.g. Polypropylene low bind, FTA DNA cards.
Quality tests (if applied): It can be analysed from evaporation rate, contaminants, fungi, dust, exogenous DNA, interaction with plastic tube, adherence or leachates, shearing.
History of users and uses: Link with item 9, 10 and 11.

### *6.3 Other Biomolecules*

Type of biomolecules: proteins, enzymes…
Extraction date:
Extraction method:
Purification method:
Preservation method: TE, water, dry, freeze-dried, other

***6.4 PCR Amplicons***

(If preserved) Reason for preserving
Location:
Amplification date:
Name (gen, fragment, etc.):
Purified: Yes/No
Fluorescence:
Amplified product size:

***6.4 DNA LIBRARY***

*This item has been thoroughly dealt with in outcome of sub-task 2.3, objective NA2*
Tissue quantity:
Date:
Method:

**7 Location:** The exact location in the collection's warehouse (room, cabinet, fridge, freezer or other). It depends on the type of specimen (tissue sample, or molecular material like nucleic acid and biomolecules) and on the preservation method (frozen, dry, liquid).

**8 Restrictions of use:** This compiles the information on the potential use of the specimen, noting if the use is free or restricted. It also includes the restriction of uses stated in the PIC and MTA agreements (Nagoya Protocol) and the ABS commitments. It furthermore includes the level of protection established by the competent legal authorities (CITES, national and international IUCN) stating permits and requirements needed for future loans.

**9 Management, record of users and uses:** This includes the information on the users, name, institution and research project. As well as the correspondence and documents (printed or digital) generated by the request. This information needs to be guaranteed by the applicable privacy law.

**10 Bibliographic records:** It collects the publications in which the specimen has been used; DOI, as well as the links to the journal web.

**11 Molecular data outputs:**

Access Number or ID of the sequence and links to any other global databases with associated information (e.g., EBI, GenBank, etc)
Name (gene, fragment, etc.):
Purification method:
Primers and/or bibliographical reference:
Amplified product size:
Sequencing primers:
Sequence consensus file (FASTA or others)

**12 Audits:** This refers to the information on: when, how and who has introduced or modified the database records.

In addition, it is also necessary to keep in mind the differences between the working database (our daily tool) and the type of information that can be made public (Web, on-line databases) which will be potentially interesting for the users (Table 1)

Table 1. Access status affecting each of the fields defined in a database

| 1 | Access information | Partially public |
|---|---|---|
| 2 | Taxonomic information | Public |
| 3 | Geographical information | Public with some exception to avoid spoliation |
| 4 | Biological information | Public |
| 5 | Information on the legal origin | Private, affected by privacy laws |
| 6 | Information on the preparation and preservation methods | Public |
| 7 | Location | Private |
| 8 | Restrictions of use | Partially public |
| 9 | Management, record of users and uses | Private, affected by privacy laws |
| 10 | Bibliographic records | Public |
| 11 | Molecular database outputs | Public |
| 12 | Audits | Private |

Finally, an important issue is quality checks. Databases need to generate periodic security backups that should be located in different venues and subject of preventive preservation. Therefore, it is advisable to establish periodic reviews of databases in order to correct or avoid typing errors, wrong nomenclature, erroneous locations, etc. This can be routine work to be done on a daily basis, although it is usually more efficient to do so once every certain time period, either by randomly checking records, or by selecting thematic blocks. Through the appropriate program scripts, it is possible to include mechanisms that would make these controls minimal.
As a case study, we present the MNCN Tissue and DNA Collection, where in order to perform a deep revision of the collection's database, the records related to the region of Madrid were selected and corrected. A detailed output of this revision can be found in Rey & Dorda, 2006.

## 4. PROTOCOLS FOR DATA COLLECTION FROM DNA EXTRACTION (DELIVERABLE 2.8)

DNA extractions should be considered with the same status as any other "specimen", and therefore, the information to be collected and linked to it should have the same treatment. Protocols for collecting data from DNA extraction should be the same as any other type of sample and should therefore follow the model described in section 3, with particular details as listed in subsection 3.6.2 and 11.
DNA extracts that do not amplify via PCR should not be thrown away, without making specific tests especially when it comes from DNA extracted from extinct, threatened, rare or old specimens. If they do not work with current amplification and sequencing techniques, the biobanks can preserve it for new techniques (GS, Hyb-Seq); this is one of the reasons why NH collections are preserved for the future: new techniques will make possible what today seems impossible.
As can be observed in the survey, only 40% of the collections retain information about the extractions, purifications and preservation methods used with DNA and this should improve. One possible reason is DNA deposits in Biobanks coming from research projects; biobanks normally lack the necessary funds to sample in the field and/or to extract DNA from such field samples, or from historical collection specimens. Generally the donated DNA is linked with basic information such as species name, date of sampling or geographical information and now, it is mandatory to carry permits, but information on extraction, preservation is in many cases missing (pers. Comm.).

## 5. CONCLUSIONS

This report shows that >90% SYNTHESYS institutions with biobanks currently provide free of charge access to their collections for non-commercial research worldwide, which is excellent. However only half the collections are actually 'discoverable' via full digitization and accessibility via external data portals, and/or links with other important databases (GenBank/GBIF/BOLD etc). Major digitization programmers are underway at many institutions, which will improve this situation, and will result in an increased demand not only for loaned research material, but in a higher quality of metadata accompanying it. Data content is currently mostly in the range of satisfactory to good for taxonomic, georeferenced and basic institutional identifier information

(although UUIDs will be necessary soon), but lacks provenance, compliance, regulatory and management (material utilization and tracking) information, as well as links to molecular analysis results and publications.
Biological information especially on specimen collected during fieldwork, preparation and preservation history (pre-analytical variables) is critical for future downstream molecular analysis by users.
This SYNTHESYS report begins to highlight the gaps here in particular, and the need for more collaboration to standardize/harmonize tissues and DNA collections.

## 6. REFERENCES


Access to Biological Collections Data task group. 2010. DNA extension for Access to Biological Collection Data (ABCDDNA), Version 2010-09-01. Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/640.

Benson, Erica E., Harding, Keith & Mackenzie-Dodds, Jacqueline. 2016. A new quality management perspective for biodiversity conservation and research: Investigating Biospecimen Reporting for Improved Study Quality (BRISQ) and the Standard PRE-analytical Code (SPREC) using Natural History Museum and culture collections as case studies. Systematics and Biodiversity, DOI: 10.1080/14772000.2016.1201167

Dessauer, H. C., Cole, C. J. & Hafner, M. S. 1990. Collection and storage of tissues. In: D.M. Hillis & C. P. Moritz, Eds. Molecular systematics. págs.25-42. Sinauer Associates Inc. Sunderland.

Dessauer, H. C. & Hafner, M. S. 1984. Collections of frozen tissues: value, management, field and laboratory procedures, and directory of existing collections. 74 págs. Association of Systematics Collections & Museum of Natural History, University of Kansas. Lawrence.

Droege, G., Barker, K., O. Seberg, J. Coddington, E. Benson, W. G. Berendsohn, B. Bunk, C. Butler, E. M. Cawsey, J. Deck, M. Döring, P. Flemons, B. Gemeinholzer, A. Güntsch, T. Hollowell, P. Kelbert, I. Kostadinov, R. Kottmann, R. T. Lawlor, C. Lyal, J. Mackenzie-Dodds, C. Meyer, D. Mulcahy, S. Y. Nussbeck, É. O'Tuama, T. Orrell, G. Petersen, T. Robertson, C. Söhngen, J. Whitacre, J. Wieczorek, P. Yilmaz, H. Zetzsche, Y. Zhang, and X. Zhou. 2016. The Global Genome Biodiversity Network (GGBN) Data Standard specification. Database 2016: baw125. doi:10.1093/database/baw125.

Hewitt, R., & Watson, P. 2013. Defining biobank. Biopreservation and biobanking, 11(5), 309-315. doi:10.1089/bio.2013.0042.

Howard, R. & Moore, A. 1994. *A complete Checklist of the Birds of the World*. 2nd ed. Academic Press. London. 630 pp.

Madroño, A., González, C. & Atienza, J. C. (Eds.). 2004. *Libro Rojo de las Aves de España*. Dirección General para la Biodiversidad-SEO. Madrid. 452 pp.

Peña, C. & Malm, T. 2012. VoSeq: A Voucher and DNA Sequence Web Application. *PLoS ONE*, 7(6): e39071. http://dx.doi.org/10.1371/journal.pone.0039071.

Pleguezuelos, J. M., Márquez, R. & Lizana, M. (Eds.). 2002. *Atlas y Libro Rojo de los Anfibios y Reptiles de España*. 2ª impresión. Dirección General de Conservación de la Naturaleza - Asociación Herpetológica Española. Madrid. 587 pp.

Prendini, L., Hanner, R. & De Salle, R. 2002. Obtaining, storing and archiving specimens and tissue samples for use in molecular studies. In: R. De Salle, G. Giribet & W. Wheeler, Eds. Techniques in molecular systematics and evolution. pp. 176-248. Birkhaeuser Verlag. Basel, Boston & Berlin.

Rey, I. 2013. Museos, colecciones científicas y ADN. Memorias de la Real Sociedad Española de Historia Natural, 11: 53-68.

Rey Fraile, I. 2014. *La conservación del patrimonio genético: colecciones de ADN y tejidos*. DoctoralThesis http://digital.csic.es/handle/10261/99240

Rey, I. & Dorda, B. A. 2006. Catálogo de las muestras de fauna de la Comunidad de Madrid conservadas en la Colección de Tejidos y ADN del MNCN. *Graellsia*, 62(número extraordinario): 175-200. http://dx.doi.org/10.3989/graellsia.2006.v62.iExtra.120.Savolainen, V. & Chase, M. W. 2006. What DNA can - and cannot - be used for. In: V. Savolainen, M.P. Powell, K. Davis, K., G. Reeves & A. Corthals Eds. DNA and tissue banking for biodiversity and conservation: theory, practice and uses. pp. 2-5. Kew Publishing. Kew.

Thomas, R. H. 1994. Analysis of DNA from natural history museum collections. Experientia Supplementum (Basel), 69: 311-321.

Sherwin, W. B. 1991. Collecting mammalian tissue and data for genetic studies. Mammal Review, 21 (1): 21-30.

Wilson, D. E. & Reeder, D. M. (Eds). 1993. *Mammal Species of the World*. 2nd edition. Smithsonian Institution Press. Washington and London. 1206 pp.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertsom, T. & Vieglais, D. 2012 Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. doi:10.1371/journal.pone.0029715

## ANNEX I SURVEY QUESTIONS

### Synthesys 3 WP2 Task 2.3
### Develop protocols for data collection from DNA extraction SURVEY

The aim of a collection is preserving and making accessible molecular genomic biodiversity to the maximum amount of researchers possible on the short, medium and long- term; including populations and species that may be extinct. Collections are also vital to preserve the orphan molecular samples resulting after the finalization of a research project.

The volume of samples managed by lab or research team is normally less than in tissues and nucleic acids collections.
From our perspective the protocols, processes and goals from both types of facilities are different.
In a lab, samples are processed to obtain specifics results from partial or complete genomes, normally linked to produce a publication. Due to this distinction the databases of both facilities can have very different contents. In the case of collections what is important is the information related to preservations, extraction methods and access for futures uses, while the techniques applied (e.g., amplification or genotyping) might not be so relevant (Although they could be considered as supplemental data). In the case of labs, due to their objective-driven nature - e.g., gender identification-, the preservation of the material may not be a priority and thus the related information (e.g, the locations) is not included in the database, however, for a collection it is imperative.

**1.- Does your institution have Tissues and DNA Collections?**

| | Yes | No |
|---|---|---|
| Free access for researchers who request loans and kept by dedicated staff | | |
| Exclusive access for staff of a specific lab | | |
| Exclusive access for a particular researcher or research team | | |

**2.- Conditions of use**

| | Yes | No |
|---|---|---|
| Free | | |
| Charge | | |
| License | | |

If LICENSE please explain e.g.http://opendatacommons.org/

**3.- Is the collection managed through a data base?**

| | Yes | No |
|---|---|---|
| Fully digitized | | |
| Mostly digitized | | |
| Other (please specify percentage if possible | | |

**4.- Regarding external Access to database**

| | Si | No |
|---|---|---|
| Analogic (including direct queries to collection staff) | | |
| Digital (accessible only in the institution) | | |
| On line (accessible from outside the institution) | | |

**5.- The database includes information on: (choose as needed)**

| | Yes | No |
|---|---|---|
| **Access** Including institutional acronym, catalogue number of the collection | | |
| **Taxonomy** From Kingdom or Phylum to species | | |
| **Geography and georreference** Locality where it was collected as exact as possible, including land or marine GPS coordinates | | |
| **Biology,** morphological, ecological data | | |
| **Origin** Type of acquisition, legal aspects ( e.g. collecting, importing or donation permits ) | | |
| **Preparation and preservation methods** | | |
| **Locations** Information on the physical location of the specimens within the collection | | |
| **Restrictions of use** | | |
| **Management, user and uses log** | | |
| **Literature outputs** | | |
| **Molecular database outputs** | | |
| **Audits** | | |

**Access information:** Including institutional acronym, catalogue number of the collection; as well as catalogue numbers from other collections that house associated specimens (e.g., voucher specimens); number of collector, and UUID (universally unique identifier) codes.
**Taxonomic information:** From Kingdom or Phylum to species, including the maximum number of subcategories possible (Tribe, subspecies, etc.). Also the information if it is a name-bearing specimen (i.e., type specimen) and/or any known synonyms, as well as the taxonomic literature used as reference.
**Geographic information:** Locality where it was collected as exact as possible, including land or marine GPS coordinates.
**Biological information:** Morphological data and/or characters like sex, age and ecological data that could be annotated during the collection and/or registration of the specimen. Also any information on anatomical or histological preparations (including images) of the specimen would be included here.
**Legal information on the origin of the samples/specimens:** This refers to the entry and origin of the specimens like the type of acquisition, collecting permits, importing permits, acceptance reports or any patronage documents.

**Information on preparation and preservation methods:** Here we would include all the data on the type of sampling applied, the transportation used, and any preliminary preservation method applied. Also any details on the kind of tissue or nucleic acid that could have been used in the acquisition process or in any posterior procedures done with the original material. In the particular case if tissue and DNA collections, normally in this field we would include the DNA extraction method applied and the final preservation technique applied to the material.
**Location information:** Information on the physical location of the specimens within the collection. The kind of information will normally depend on the kind of specimen (e.g., tissue sample, nucleic acid) and the preservation method applied (freezing, dry preservation, preservation in fluids).
**Restriction of use:** Data on the possible uses of the specimen indicating if there are any restrictions and/or delays when loaning or providing access.
**Information on management, uses, and/or user records:** Information on who has used the specimens, name of their institutions and research projects. Also any messages and documents (printed and digitized) related to the access/loan request.
**Literature outputs:** Compilation of scientific publications that have used the specimen
**Molecular database outputs**: Links to any other global databases with associated information (e.g., EBI, GenBank)
**Audits:** Information on when, who and how a record has been added or modified.

**6.- Regarding acid nucleic samples, your collection retains information about the extractions, purifications and preservation methods used?**

| Yes | No |
|---|---|
| | |

**7.- Regarding tissue samples, your collection retains information about the preparation and preservation methods used?**

| Yes | No |
|---|---|
| | |

**8.- Can you select any methods used on next possibilities?**

| | Yes |
|---|---|
| Phenolchloroform extraction | |
| Phosphate buffered saline DNA extraction | |
| Silica based DNA extraction method | |
| Chelating agent extraction | |
| Commercial kit DNA extraction | |

Other (can you write any of them?)

**ANNEX II. Partners invited to participate in the survey**

| | | |
|---|---|---|
| 1 | NATURAL HISTORY MUSEUM | NHM |
| 2 | ROYAL BOTANIC GARDENS KEW | RBGK |
| 3 | ROYAL BOTANIC GARDEN EDINBURGH | RBGE |
| 4 | MUSEUM NATIONAL D'HISTOIRE NATURELLE | MNHN |
| 5 | MUSEO NACIONAL DE CIENCIAS NATURALES (CSIC) | CSIC |
| 6 | REAL JARDÍN BOTÁNICO (CSIC) | CSIC |
| 7 | NATURHISTORISKA RIKSMUSEET | NRM |
| 8 | STICHTING NATURALIS BIODIVERSITY CENTER | NCB |
| 9 | MUSEUM FUR NATURKUNDE - LEIBNIZ-INSTITUT FUR EVOLUTIONS- UND BIODIVERSITATSFORSCHUNG AN DER HUMBOLDT-UNIVERSITAT ZU BERLIN | MfN |
| 10 | NATURHISTORISCHES MUSEUM | NHMW |
| 11 | INSTITUT ROYAL DES SCIENCES NATURELLES DE BELGIQUE | RBINS |
| 12 | MUSEE ROYAL DE L'AFRIQUE CENTRALE | MRAC |
| 13 | NARODNI MUZEUM-NATIONAL MUSEUM NM | NMP |
| 14 | MUSEUM OF ZOOLOGY SENCKENBERG | |