African BioGenome Project

Genomics for the future of biological diversity across Africa

Sample Collection and Processing Sub-committee

Terms of Reference: V0.9

Contents

Audience

This document provides a summary and outline of the main terms of reference (ToR) and standards of the *Sample Collection and Processing Sub-committee* of the African BioGenome Pilot Project. The document provides broad guidance around ethical collecting, metadata standards and how to preserve and process samples and, should be referenced whenever a sample is intended to be submitted to the project. The document contains specifications for a large range of phyla and taxa and as such may need some ongoing improvement. It is likely to grow, be refined or changed completely as we progress with the Pilot Project and obtain greater knowledge and experience with various taxonomic groups. The Terms of Reference will thus gradually be refined and host increasing information and resources as the project matures. Should you wish to contribute to these guidelines, please contact the Sample Collection and Processing Sub-committee here.

Phases

The Project aims to sequence a representative of every Eukaryotic species on the African continent ranging from terrestrial to marine species. As this is an incredibly ambitious endeavour, this will be completed in a phased approach. These phases have been briefly outlined below.

Pilot Phase: The pilot project phase will aim to sequence up to 2500 endemic African species and is planned to run from November 2021 until March 2024. These species will be sourced from across Africa and the committee is currently collecting nominations from the broader community for species to be included and prioritised as part of this phase. The goal is to distribute sequencing efforts across regional diversity and taxa and the sub-committee will work closely with established taxonomic working groups (housed within the Sample Collection and Processing Sub-committee) to nominate and prioritise representative species. One of the first key tasks within this pilot phase, is to sequence an initial 18 vertebrate species in collaboration with the Vertebrate Genome Project (VGP) [https://vertebrategenomesproject.org/] as well as a further 100 plant species in collaboration with the 10 000 Plant Genomes Project (10KP) [https://db.cngb.org/10kp/]. These species will be sourced from various African countries, sequenced locally (within Africa) and the data then contributed to the efforts of the VGP and the 10KP project. These species will also serve as a “dry-run” to optimize processes and workflows from sample collection to sequencing and analysis.

Phase 1: This phase is planned to run from March 2024 until April 2028 and will act as the large-scale implementation of the project. We broadly aim to sequence the following:

- Up to 45 thousand plant species
- >60 thousand animal species
    - 50 thousand invertebrates
    - 3360 fish
    - 800 amphibians
    - 2026 reptiles

- 2341 Birds
- 100 mammals

This phase is likely to be further subdivided into additional sub-phases as follows:

- Phase 1a - sequencing a representative of every family
- Phase 1b - sequencing a representative of every genus
- Phase 1c - sequencing a representative of every species

Phase 2: The second phase of the project will run from April 2028 - March 2032 and will largely be used to re-assess and re-evaluate efforts from Phase 1. This phase will be used to sequence any additional taxa not covered during the Pilot Phase and Phase 1.

Considerations for selecting representative species:

Community/economic value - we suggest sequencing species with broad economic/ community impact and value. Species that loosely meet one or more of the following criteria will be viewed as a priority:

- ☐ Agriculturally important species
- ☐ Human and veterinary diseases organisms / vectors / intermediate hosts
- ☐ Species of forensic interests
- ☐ Quarantine species
- ☐ Species of economic impact
- ☐ Indicator and sentinel species
- ☐ Species that damage or pose risks to human infrastructure / activities
- ☐ Circumtropical species with wide distribution
- ☐ Species that have little information about them in literature

Taxonomic status - the species contributed to the project should be under no taxonomic disagreement and should be an accepted, preferably well described species.

Genome size - taxa with smaller diploid genomes will likely be prioritised due to the reduced financial and human capital required to sequence and assemble these genomes. However, other species will still be considered where resources allow.

Physical size - as we are sourcing species from all over Africa, where possible, smaller organisms may be prioritised in cases where the whole organism will be shipped for further processing. Where blood/tissue/other samples are submitted, this will be overlooked.

Country of collection - we aim to involve as many countries in the project as possible however during the pilot phase, we will prioritise samples with less legal overhead and where established relations may already exist between the source country and the country where sequencing will be performed. We will also prioritise countries where we are able to obtain the relevant permits and legal documentation initially but will work to expand our efforts to more regions as we progress. In order to build the capacity of African institutions and scientists, priority will be given to local institutions with inhouse sequencing platforms. During the pilot phase this capacity will be assessed and a database of all existing sequencing platforms created.

Ethical collecting

The African BioGenome Project will strictly adhere to the Nagoya protocol [
] as has been suggested by other similar efforts (see https://www.earthbiogenome.org/sample-collection-processing-standards), particularly since most samples are likely to be shipped out of their home country to another African country for further processing. In cases where samples will be moved outside of their home country for sequencing, the rules around 'Access Benefit Sharing', suggested by the Nagoya protocol, must strictly be adhered to. You can find out more about Access Benefit Sharing and how to get in touch with your ABS clearing house here: https://www.cbd.int/abs/theabsch.shtml. Two vital documents which the Nagoya protocol requires in order to move specimens out of their home country are a Prior Informed Consent (PIC) as well as Mutually Agreed Terms (MAT) document that outlines what the benefit of sharing the sample will be e.g. authorship on a paper, long term access to data generated etc. AfricaBP has developed a 'Material Transfer Agreement (MTA)' to facilitate this process. MTA's will be developed per species submitted to the project and all relevant stakeholders will need to agree on these terms ahead of proceeding with sequencing.These agreements must ensure that countries receiving specimens are able to perform the analyses required but must also ensure the specimen may be shipped to an additional country/ies should the project require it.

It is of utmost importance that any sample/specimen/tissue contributed to the African BioGenome Project be sourced ethically and that all sample custodians/ambassadors (collectors and contributors of samples) adhere to institutional, regional and national guidelines and frameworks (appropriate permissions, consent, permits, approved ethical clearance, Nagoya protocol, etc.). It is thus vital that all necessary permits, documentation and approvals are in place ahead of performing any sample collection and sequencing, and formally contributing a specimen to the project. The African BioGenome Project will require a copy of all such permissions including any ethical clearance documentation for each sample submitted to the project (please see checklist for a comprehensive list of the documentation you will require or see summary below). These files will not be made public but will be kept as surety that specimens have been sourced legally and ethically should any evidence be required. Copies of all such documentation will be requested before formally agreeing to include a species in the project. All ethical and legal documentation will be stored and managed by the Ethical Legal and Social Issues Sub-committee within AfricaBP but this process will be facilitated by the Sample Collection and Processing Sub-committee.

List of documentation to be provided with every sample (checklist):

1. Collecting Permit
2. Export Permit
3. Import Permit
4. Mutually Agreed Terms (MAT)
5. Prior Informed Consent (PIC) - if required
6. Letter informing Country of Origin of 3rd Party Transfer
7. Internationally Recognised Certificate of Compliance Number/s
8. CITES Registration code of supplier Other (please specify e.g. hazard statement; conservation or preservation history
9. Sample Metadata File

National representatives have been appointed across many African countries who possess comprehensive prior knowledge of national collection standards and best practices (typically an Access Benefit and Sharing focal person within a particular country) and they will assist sample custodians/ambassadors to prepare the required documentation and assist in acquiring the appropriate permits, clearance etc. All sample custodians must work with their national representative/s (should one be appointed for their country). In addition to national representatives, regional coordinators have also been appointed to work with both national representatives and sample custodians/ambassadors to facilitate the sampling process and will provide additional oversight to collections processes.

In nearly all cases (this will be taxa specific), we will also require sample ambassadors to prepare additional tissue to be submitted for barcoding (please read the metadata section below for more information), should the sample custodian/ambassador not be in a position to perform and provide this barcode themselves. When a custodian is able to and possesses the necessary expertise, an RNA sample collection may also be performed and provided along with the main sample.

In all cases, a wild, ethically-caught specimen is preferred over a laboratory or other similarly captive/cultivated/cultured specimen but specimens from zoos, museums, and similar may be accepted. Where traditional knowledge has been used in order to gain access to or source a sample, this contribution must be explicitly recorded in the sample metadata and included when submitting the sample.

Collecting metadata

In order to ensure accuracy and good sample/data provenance, all species submitted to the project must be accompanied by a standard sample metadata file, voucher specimen as well as a DNA barcode (we will assist should the custodian not have the required expertise). The voucher specimen can take many forms including a blood or tissue sample (or culture for Fungi and Protists), DNA or RNA, but must accompany the specimen where possible. The DNA barcode must also be submitted to a public repository (we suggest using BOLD or GENBANK) once generated.

Detailed photographs should accompany all specimens particularly where a voucher may not be available (e.g. in the case of invertebrates where the whole animal may have been used for downstream processes). These images will be made publicly available and so images are preferred that the custodian has taken themselves or obtained appropriate permissions for (in writing). A consent to store and make images publicly available must also be completed and submitted in the event that images taken by another individual are submitted by a sample ambassador/custodian. The appropriate copyright rules will apply.

All species must be taxonomically identified to the species level by a taxonomist with the required expertise and skills (the identifier's information will also be captured). The African BioGenome Project has a number of taxonomic experts within their community and should you require their assistance, a request can be made to africabp-core-members-sample-collection-processing@googlegroups.com. Species must be taxonomically accepted and not be undergoing any disputes or confusion.

Along with accurate taxonomic information, a range of metadata must be collected and submitted along with the sample/s. Details requested as part of the metadata will include, collector's information, identifier's information, specimen taxonomic information, sample information as well as fields collecting locational data and/or environmental information, etc. We have outlined what is considered minimal (i.e. information we must receive to process your sample) and what is considered optional (good information to have for scientific reasons) within our metadata collection sheet. All minimum data fields must be completed to process a sample. This metadata sheet can be accessed by clicking here or requested from the aforementioned email address.

Specimen/tissue preservation and processing

Freshly collected and subsequently flash frozen (with liquid nitrogen) and stored at -80 specimens/tissue are preferred over specimens/tissue stored in any other preservation media. As indicated by the VGP, flash freezing is currently considered the "gold standard" for high molecular weight DNA extractions. Dying/dead specimens may already have experienced DNA and tissue degradation which would result in lower quality DNA downstream and must therefore be avoided whenever a fresh alternative is available. While there may be exceptions, material collected from museums and herbaria will typically not be sufficient for high throughput sequencing. In instances where a living specimen cannot be collected/sampled, time taken to preserve the sample and the storage media used becomes increasingly important.

Processing standards will vary greatly depending on the specific taxonomic group and it will be impossible to cover all specifications for each group here, however we have made some broad specifications available below:

Animals (vertebrates and invertebrates), Protists and Fungi
In keeping with recommendations from other similar projects, AfricaBP will require roughly 10 - 100mg of tissue per 1Gb of genome size.
The GOAT project (https://goat.genomehubs.org/) can assist in determining your likely genome size by comparing your species to sister taxa within the system. Useful when determining how much tissue to submit. This tissue volume must be submitted per technology type used e.g. Hi-C, PacBio, etc. As AfricaBP may perform PacBio, Bionano and Hi-C sequencing, we require ~60 - 200 mg of tissue (genome size dependent) and an additional 50 - 100 mg to account for any additional sequencing that may need to be performed. A smaller aliquot of ~10 - 20 mg must be provided should barcoding need to be performed by AfricaBP. The 200mg of tissue (or more) must be divided into vials each containing > 20 mg tissue to avoid frequent freeze-thaw cycles. The best tissue type to use will differ across species but good tissue types to use include: muscle (>50 mg prefered) , liver, spleen, kidney, brain and gonads.

For Fungi, where possible, fresh cultured mycelium would be the preferred option. For many macrofungi, this will not be possible and fruiting bodies will need to be sent in appropriate formats/preservatives.

For Protists, some will be able to be cultured but many will exist as single-celled, free-living organisms (many will be aquatic). These have their own particular problems/issues and require careful handling (not least as many carry their own endosymbionts).

Plants

For plants and Chromista the European Biogenome Project suggests at least 100 and preferably 1000 milligrams of tissue per Gb of genome for each sample. Young leaves, dark treated for 24-72hrs are preferred. The leaves should be healthy, no dead tissue and no stems. Plants and Chromista should also be processed into small pieces to support rapid freezing, but larger volumes of tissue might be placed in tubes given up to ten times more tissue may be required to achieve adequate quantities of DNA for these groups.

The above protocols broadly apply however specific volumes, preferred tissue types, etc. are likely to change dramatically across taxa. A range of PacBio specific protocols and publications for various species and phyla have been curated by PacBio here: https://extractdnaforpacbio.com/. Please use these publications as a guide for sample collection and processing. In time, this committee will also produce a set of taxa specific SOP's which we will make freely available to the public.

Sample submission:

Prior to shipping the sample, an AfriaBP specific project barcode/UID should be requested from the sampling committee and assigned to each species/specimen submitted. Ahead of freezing and shipping it is preferable for the specimen to be photographed with barcode/s in view to assist in resolving any potential sample confusion that may arise downstream. A single species may require multiple aliquots which must be easily linked back to the main species/primary sample. The specimen should thus also be photographed alongside any vials, tubes, etc. that may be used to ship the specimen tissue aliquots, with tube barcodes visible as well. Each tube containing materials to be shipped must be captured in the metadata sheet with tube (e.g. flexicodes) recorded and captured in the "sample photo". Should any vials be mixed up during shipping, this should allow us to easily correct these missteps.

Once processed/frozen, samples must be stored at -80℃ and shipped to the sequencing country on dry ice. The sample must under no conditions thaw during this time and the appropriate packing etc. must be used to ensure a viable specimen reaches the sequencing facility. Where possible, sample ambassadors should try to use enough dry ice to last the entire period of transit. If possible, shipment on a Friday should be avoided to prevent transit over the weekends. Some shipping companies may provide dry ice top ups in case of delays as well. Often the optimum solution is to ship tissue overnight however this may not always be possible in all African countries thus trusted couriers and partners should be prioritised. Please see shipping instructions for more information.

Processing samples without liquid nitrogen:

While flash freezing and overnight shipping may be the gold standard for high throughput genomics, this will likely not be a viable option for many African researchers. Where liquid nitrogen is not available for flash freezing, a combination of dry ice and ethanol may be used for tissue collection and storage. Freshly collected tissue may be immediately stored within at least 96% ethanol, but preferably 99 - 100% ethanol and placed at -20℃ as soon as possible after collecting, or preferably, placed on dry ice immediately until the sample can be

placed into a deepfreeze. Where possible, the tubes containing ethanol for preservation may be placed on dry ice to cool for about 5-10 minutes before inserting the freshly excised tissue into this cooled ethanol and subsequently stored at -20℃ until shipping. The tissue must then be shipped on dry ice as soon as possible after collecting. Specimens must reach sequencing facilities no more than a maximum of 3 days after shipping (couriered overnight whenever possible). Should the specimen be degraded upon arrival at the sequencing facility, the specimen will not be processed and replacement tissue will be requested. Please note that this method must be used only in cases where flash freezing with liquid nitrogen is not an available option and additional taxa specific steps may be required to ensure DNA integrity - for example, with gelatinous marine tissue that typically contains a very high water weight, an ethanol replacement may need to be performed every 12 hours after collection for the first 24 hours to ensure ethanol concentration and tissue integrity is maintained.

Detailed protocols for various tissue types will become available as the project matures.

Specimen/tissue shipping

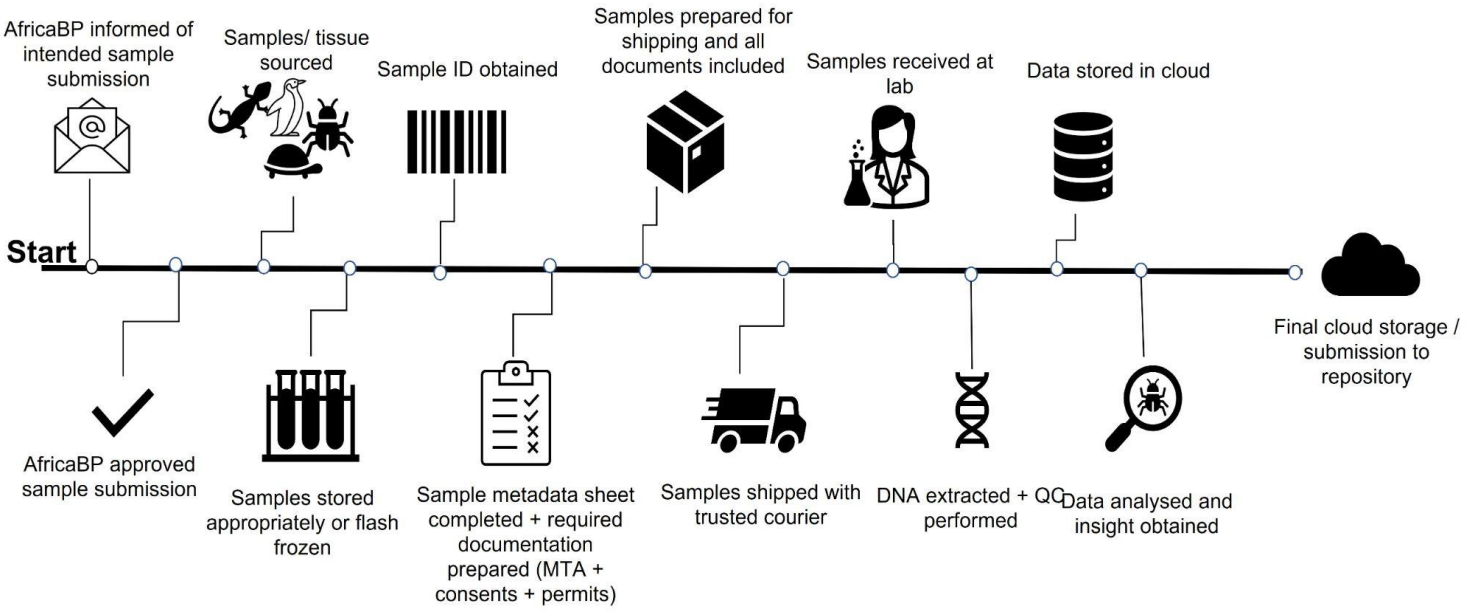
Assistance in obtaining the appropriate shipping labels and clearance documentation can be sought from your national representative or the sample collection and processing committee by lodging a request here: africabp-core-members-sample-collection-processing@googlegroups.com. We will not be held liable for “lost” samples so ensure shipping labels, addresses and any additional permits required etc. are sourced well ahead of time. Please also consider sample degradation and use appropriate storage mechanisms as outlined under “Tissue Preservation” to ensure viable tissue reaches the sequencing country.

When you are ready to ship your sample, please ensure the following checklist has been met:

- ☐ AfricaBP has previously approved your specimen/sample contribution
- ☐ Legal, ethical and other vital information regarding the specimen collection has been submitted to AfricaBP i.e. Mutually Agreed Terms, Consent, Collection Permits, etc.
- ☐ Relevant shipping labels, containers and documentation have been acquired
- ☐ Samples have been photographed with all tube codes in view
- ☐ Barcode has been provided to AfricaBP OR tissue for barcoding included in sample to be submitted
- ☐ Metadata sheet has been completed and a copy included with the sample as well as sent to AfricaBP for verification
- ☐ Samples + relevant documentation have been included in and packaged appropriately ahead of shipping

You may use any courier service to deliver your samples however AfricaBP suggests using either UPS, FEDEX or DHL courier’s services as these couriers have proven experience in transporting biological material, have a range of shipping materials available and are well placed to assist with country specific shipping issues.

Graphical overview of sampling and processing pipeline



Contributors to committee Terms of Reference

Ms. Verena Ras (Committee Chair)
H3ABioNet, Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, CIDRI Africa Wellcome Trust Centre, University of Cape Town, South Africa.
verena.ras@uct.ac.za

Dr. Alan Buddie
Center for Agriculture and Bioscience International (CABI), Bakeham Lane, Egham, Surrey, TW20 9TY, United Kingdom
a.buddie@cabi.org

Mr. Joseph Kirangwa
University of Cologne, Worm-Laboratory [worm-lab.eu], Zülpicher Str. 47b, 50674 Köln, Germany
jkirangw@uni-koeln.de

Dr. Katali Kirungi Benda
National Animal Genetics Resources Centre and Data Bank (NAGRC&DB), Plot 98 - 106 Nsamizi Road, P.O. Box 183, Entebbe, Uganda.
katalibenda@gmail.com