GAGA Sample Collection Guidelines and Explanations
June 2018

To be useful for cross-species comparisons, sample collection for GAGA should follow the recommendations outlined and explained here both for the minimally required and the ideal samples to be used for DNA and RNA extraction. Please always use the form (or excel sheet) at http://antgenomics.dk/gaga-submission/ when collecting and make sure to collect additional data (habitat, life-history information, etc. wherever possible) and material (vouchers) for each collected colony.

Minimum requirements

The minimum requirements for high-quality reference genomes are sufficient sample material for extracting 12 µg of genomic DNA and 2 x 1 µg of RNA (1 µg from workers and 1 µg from gynes/queens).

This approximately translates into ca. 120 small ants (~ 2 mm body length) or 12 large ants (~ 10 mm body length).

In addition, it is crucial to collect morphological voucher specimens in ethanol and genomic voucher specimens in RNAlater.

Similarly, we ask you to submit information about each collected colony using the sample submission form online (http://antgenomics.dk/gaga-submission/) or the excel sheet available online (http://antgenomics.dk/wp-content/uploads/GAGA_rawSampleSubmission.xlsx).

If possible, collecting more material than what is minimally required is highly recommended.

Samples for reference genome sequencing may consist of both adults and brood and should be taken from the same colony.

Detailed descriptions and additional recommendations are below.

Genomic/transcriptomic samples should be stored in RNAlater.

The use of RNAlater is very straight-forward and does not require wet-lab experience. RNAlater deactivates nucleases for 24 hours at 37°C, 7 days at 18-25°C, 2 weeks at 4°C and indefinitely at -20°C. See below for detailed instructions.

Voucher collection

For every target species **whole-nest series have to be retained for museum and genomic collection vouchers**. Whole-nest series comprise specimens of brood, workers (including morphologically differentiated subcastes), reproductives (males and gynes) and whenever feasible mother queens and closely associated symbionts (e.g. fungus garden or aphid samples). One series each will be submitted to archival genomic collections (e.g., the Smithsonian NMNH Biorepository) and taxonomic museum collections. Contact Morten Schiøtt (MSchiott@bio.ku.dk) or Jon Shik (jonathan.shik@bio.ku.dk) to organize specimen deposition. Museum vouchers should be stored in ethanol, while genomic vouchers should be stored in RNAlater or snap frozen and stored at -80° C (see above). Morphological vouchers will furthermore be used for micro-CT scans that we intend to perform for workers and sexuals of each GAGA species. Micro-CT scans will not damage the voucher specimens.

Detailed genome/transcriptome sampling recommendations

The following guidelines specify the ideal set of samples to generate the most comprehensive datasets for the ant species targeted by GAGA. In mass, the minimum requirements for high-quality reference genome are sufficient sample material for obtaining 12 µg of DNA and 1 µg of RNA from gynes or queens and 1 µg of RNA from workers.

1. Samples for reference genome sequencing and assembly

Samples for reference genome sequencing may consist of both adults and brood and should ideally be taken from the same colony. DNA quality (i.e. level of degradation) largely depends on sampling conditions (see also below). Reference genomes will be generated by a combined approach using PacBio long-reads and Illumina short reads. Most of the genomic DNA is required PacBio libraries, which are crucial for high quality (long contigs) genome assemblies.

Summary: *Minimum requirement for obtaining an ant reference genome: Sufficient biomass to extract 12 µg of hiqh-quality genomic DNA.*

2. Samples for transcriptome sequencing

To annotate the genome and assess caste-specific gene expression it is highly advantageous to have sequenced transcriptomes of different castes and life-stages. Transcriptome samples should preferably be obtained from the same colony as the samples to be used for obtaining the sequenced reference genome. To cover as much gene-expression diversity as possible, samples of brood (eggs, larvae and pupae, ideally separated by sex and caste), different worker phenotypes (adult sub-castes, callows, foragers, nurses, etc.), and reproductives (males, queens, gynes) should ideally be sampled. **Collection of female sexuals and workers is a minimum requirement for any species. Samples**

from different castes and life stages should be stored in separate vials to retain the possibility of acquiring specific transcriptomes. For each transcriptome at least 1 µg of total RNA is required.

Summary: *Minimum requirement for reference genomes: Sufficient biomass to extract 1 µg of total RNA from workers and 1 µg of total RNA from gynes or queens.*

3. Samples for metagenomics of ant-associated microbial communities

Whenever feasible, worker samples from three colonies from different geographic locations should be collected. If this is not fesasible, we ask you to collect workers from up to three different colonies from the same geographic location. These will be used for 16S DNA sequencing to characterize the typical microbiome of each ant species. Data from different geographic locations will allow the differentiation of species-specific and population-specific symbionts. **Please sample ≥ 10 workers in RNAlater or ethanol (regardless of individual size). These workers should not be mashed, crushed or cut into pieces.**

Summary: *Minimum requirement for microbiome analyses: 3 x 10 workers from different colonies, stored in RNAlater/ethanol (not mashed or cut in pieces).*

4. Samples for parent-progeny sequencing

Whenever feasible, the mother queen and several male offspring of the same colony should be sampled for parent-progeny re-sequencing studies.

Sample data collection

The following information about collected colonies is minimally required for all samples submitted to GAGA. Please use the online sample submission form (http://antgenomics.dk/gaga-submission) or the excel file () to provide the following data for each colony (i.e. each "collection event").

1. **Collection data:**
   Collector name, collection date, collection ID.
2. **Sampling site data:**
   Collection coordinates (latitude, longitude and elevation), nesting site/microhabitat (arboreal, subterranean, etc.), habitat (primary forest, road side, etc.).
3. **Life history data:**
   Approximate number of queens, approximate colony size.

If possible, please take a photograph of the collection site (facing north).

Labelling & record keeping

We kindly ask you to use 2 ml screw-cap collection tubes whenever possible to facilitate storage on our end. Please also label each tube individually with the unique collection ID assigned by the collector. For RNAlater-stored samples, please label the outside of the tube and **do not** add a tag to the tube.

Please make sure to keep records **for each tube** (ideally in an excel or google spreadsheet) so that it is clear from these records what species and what type of sample (e.g. "10 workers") is in each tube you submit. This is crucial as a sample stored in RNAlater cannot be screened for its contents when frozen. When sending samples to CSE, please make sure to also send the file containing information about all tubes by email to Morten Schiøtt (MSchiott@bio.ku.dk), Jon Shik (jonathan.shik@bio.ku.dk), or Lukas Schrader (Lukas.Schrader@bio.ku.dk).

Shipping

All samples should be **shipped on dry ice**. Normally around 10 kg of dry ice will suffice, but no risks with marginal quantities of dry ice are to be taken at this phase. Shipments should be submitted on Mondays to prevent arrival in Copenhagen on weekends. Please coordinate with Morten Schiøtt at CSE (MSchiott@bio.ku.dk) regarding shipments.

Permits

In general, we will only be able to use material collected in agreement with the Nagoya protocol of 2010 when collecting in countries that have signed that treaty. This means you need to secure you are entitled to collect. This is often the case for nationals or citizens of Nagoya countries, either in general or in particular when collection sites are part of nature reserves as long as collectors hold national permits. The full information is available at www.cbd.int/abs/.

A permit to import biological samples to Denmark will be needed. Before sending the samples, contact Morten Schiøtt, who will submit an application for the permit. Permits to export samples will have to be held by the sender.

DNA/RNA yield of ants

A rough estimate of yield is **1-2 µg DNA/RNA per 10 mg of adult ant tissue.**

Based on previous experience, 10 µg of genomic DNA requires about 100 workers of very small ants (~2 mm body length, e.g. Monomorium pharaonis) or 10 workers of larger ants (~10 mm body length, e.g. major workers of *Acromyrmex echinatior*).

1 µg RNA would require ca. 10 workers of small ants or two workers of larger ants.

Sample preservation

DNA quality (i.e. level of degradation) largely depends on sampling conditions and preservation. Highly degraded DNA cannot be used to generate large-insert libraries that are crucial for the generation of high-quality assemblies. Likewise, degraded RNA cannot be used for transcriptome sequencing.

Sampling in RNAlater

While RNAlater is not toxic, we suggest wearing plastic gloves throughout the protocol.
Start by preparing the required number of sampling vials by filling them to 2/3 with RNAlater.
Please make sure to sterilize forceps/pestils between different samples. Place the samples in 5–10 volumes of RNAlater Solution. Use RNAlater with fresh samples only. Do not freeze ants before immersion in RNAlater. It can be useful to chill live ants at 4° C to immobilize them prior to sampling.

Based on our experience with extracting DNA/RNA from RNAlater stored samples, crushing or squishing ants does not improve quantity and quality of the DNA/RNA. We thus no longer ask collectors to crush individuals stored in RNAlater.

Sampling in RNAlater can be difficult as the ants do not die as quickly as in ethanol. It can help to chill the ants in a fridge before sampling.

Store all the RNAlater samples at **4° C overnight** and transfer them to **-20° C or -80° C afterwards**. It is crucial to store the samples overnight at 4° C first before storing them at lower temperatures to prevent precipitation of salts.

As an alternative to RNAlater, but likely less feasible in the field and for shipping, samples can be snap-frozen and stored at -80° C (or colder) WITHOUT ETHANOL having been added. Such samples can also be transferred to "RNAlater ICE" for shipping, to make them less sensitive to suboptimal temperatures. As a last resort, DNA (but not RNA) samples can be stored in 95-100% ethanol, but this will produce very fragmented genomes, which will be of limited value for comparative analyses across the ant tree of life. Any such ethanol samples should be placed into -80° C or colder storage as quickly as possible.

Below you can find some more notes regarding sample collection that we developed over the last couple of months.

The best possible multipurpose sample for GAGA has:

- Workers – preferably from ≥2 sites, so we can capture some of the geographic variation in gut-microbiomes
- Gynes from at least one colony for the transcriptome
- Males from at least one colony for the transcriptome
- Brood from at least one colony for the transcriptome
- Mother queen and her sons – this will allow comparative analysis of mutation rate - often hard to get but worth the effort if feasible
- Additional workers from 1-2 other (remote) sites so we can appreciate how spatially variable the gut microbiomes are (possibly tricky when cryptic species exist)

This implies that:

1. The ideal GAGA sample consists of one large colony that has everything and with enough biomass: workers, gynes, males, brood, and the mother queen(s).

2. 2-3 other colonies of the same species from the same site, for having replication for gut microbiota and for increasing the likelihood of getting at least one mother-queen associated with male offspring.

3. Although this may be difficult in the tropics, it is very important to avoid sampling several cryptic species under the same name – hence getting everything from one big colony is preferable.

4. In practice all this means: when possible collect a series of nests in boxes that keep colonies alive for a day, and then decide in your hotel or field station what is the best primary colony and the optimal satellite colonies to be preserved.

5. Remember to also collect some individuals in ethanol as voucher specimens, ideally for each colony separately.

Of course this does not mean that any "worker-only" sample or a very rare singleton which represents a phylogenetically critical taxon should not be secured unreservedly if possible as a colony and even as a single individual (e.g. *Tatuidris*; *Martialis*). But the crux is that most of the 200 GAGA genomes should ideally meet the wish list above (we know mother-son combinations and multiple worker samples from several distant sites will possibly be achievable only for <100 ant species).

Some general notes for working with RNAlater

- Transfer animals to screw cap tubes with RNAlater (with a volume ratio of RNAlater/tissue of maximally 5).
- ~~To ensure proper penetration of RNAlater, mash the animals lightly with clean forceps or something similar. For microbiome samples, please leave the animals intact and unmashed~~
- Best is to leave samples at 4 C for max 24 hours, i.e. essentially freeze them quickly as possible at -80. However, if that is not feasible (field conditions) samples can be left at ambient temperatures for some days or at 4 C for up to a month.
- Remember, that for large colonies, it is more important to get a small amount of properly preserved tissue, rather than a large amount of poorly preserved tissue. About 12 large ants (10 mm) or 120 small ants (2mm) will be needed for extraction of genomic DNA, and on top of that we need small amounts of as many castes and developmental stages as possible for RNA extraction.

RNAlater FAQ

If I collect into RNAlater, I need to crush or break open the ants, correct?

- ~~Yes, at least for larger ants (> 5mm). I tend to either separate the head and the gaster or squish and soak the ants while they are submerged in RNAlater~~.
- We no longer recommend crushing the ants.

If I collect into RNAlater, I need to keep it cold, correct? If, during time periods when I do not have access to refrigeration, I would need to put the RNAlater specimens in a cooler with some ice packs. Is that OK?

- Simply put, its always better to store the samples cold. According to the RNAlater manual, samples can be stored at 25°C for one week, at 4°C for one month, or at -20°C indefinitely. However, since we want high quality RNA/DNA, I would start to get nervous when storing the samples at room temperature for over a day or two. In general, samples should be kept at 4 C over night and then transferred to -80 or as cold as possible.

Do you need multiple tubes containing only one individual (gyne, male, worker)?

- This is not really necessary, but in cases where you are not limited by tube number, it will maximize the usability of samples for different purposes. Sampling individuals could allow for an entirely different range of studies (e.g. more detailed transcriptomics, population genomics, etc.), which are not planned yet, but who knows what will come up. For the scope of GAGA, we don't really need the individuals kept separately (except for the queen and her male offspring, but these samples are even harder to get anyway). So bottom line, unless the samples are really unique and rare, I don't see a necessity to keep them individually.

Should I try to separate out the biomass necessary for a genome and put it into one or more tubes of RNAlater, crushed up with a sterile wooden stick or pipette tip?

- Yes. That is possible. We can also work with multiple separate tubes for the genome. Just make sure they come from the same colony.

Should I create additional separate vials, each containing multiple individuals of each caste (males, gynes, when present) and perhaps another set containing brood (eggs, larvae, pupae)? If a queen is present, should I try to put her in her own vial of RNAlater?

- Yes. That would be exactly what we hope to get.