GAGA Sample Collection Guidelines and Explanations
June 2017

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. Always use the standard form (antgenomics.dk) when collecting to make sure you also collect additional data (habitat, life-history information, etc. wherever possible) and material (vouchers) for each focal ant species concerned.

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 (from workers and 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 necessary to collect morphological voucher specimens in ethanol and genomic voucher specimens in RNAlater.

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 Sean Brady and/or Ted Schulz 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).

Detailed genome/transcriptome sampling recommendations

The following guidelines specify the ideal set of samples to generate the most comprehensive datasets for optimal genome assembly for the ant species targeted by GAGA. In mass, the minimum requirements for such a 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 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. **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. 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. Sufficient biomass to extract 1 µg of genomic DNA will be required (10 small ants/1 large ant).

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.

5. Samples for population genomics
Whenever feasible, specific samples (e.g. workers, gynes) from 5-10 independent colonies from the same population should be collected. Such samples allow for assessing the genetic diversity within a given population by low coverage (e.g. 5x) re-sequencing of several individuals from a few colonies. Population genomic re-sequencing is not required for reference genome assemblies and is currently not part of the standard genomic pipeline of GAGA. However, having samples available for the targeted species could prove invaluable for studying population diversity.

Life history data collection

For each collected colony, as many additional data as possible should be collected. The following range of data is minimally required:

1. **Collection data:**
Collector name, collection date and time, storage conditions (i.e. has the colony been kept alive after collecting)
2. **Sampling site data:**
GPS coordinates, weather, temperature, nesting site (arboreal, subterranean, etc.), habitat (primary forest, road side, etc.)
3. **Life history data:**
Approximate number of queens found in a colony, presence of sexual offspring (larvae, pupae or adult) and symbionts, approximate number of nest chambers, colony size in order of magnitude of a typical mature colony of the focal species (10^2,10^3,etc.), approximate population density (rare species, uncommon species, abundant species), average worker size (in mm body length).

Summary: *Please use the form downloadable from the GAGA website (antgenomics.dk) to record life history data.*

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.

For larger ants (> 5 mm body length) it is necessary to crush the cuticle to allow the RNAlater solution to penetrate and soak the tissue. **T**ake a single specimen with forceps and submerge it in a tube filled with RNAlater. Subsequently keep the individual submerged and start squishing the cuticle with hard forceps or a pestle, so that the inner tissue gets soaked with RNAlater. In addition, it can improve preservation if the sample is cut into smaller pieces. According to the manufacturer's protocol, samples stored in RNAlater should be smaller than 5 mm^3 to guarantee full penetration.

For smaller ants, you can transfer several to many specimens in an RNAlater filled tube at the same time. Again use the forceps or a pestle to squish as many individuals as possible, so that the ants get soaked with RNAlater.

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.