

**Anopheles\_Metadata\_Manifest\_V4.0 SOP**

Published Date: JULY 2022

| ***Version*** | ***Date*** | ***Changes*** | ***Contributors*** |
| --- | --- | --- | --- |
| **4.0** | JULY 2022 | based on Darwin Tree of Life manifest but extensively revised to support mosquito collections. | Mara Lawniczak, Marilou Boddé, Petra Korlevic, Katie Woodcock, Vickie Brooks, Lyndall Pereira, Alex Makunin, and members of the Darwin Tree of Life Sample Working Group |

## 

## Correct and comprehensive recording of sample metadata is critical to the long-term utility of the work we do. Metadata will link the sequence data we generate to their origins. Please read this SOP in full before completing the Anopheles\_Metadata\_Manifest.

## Please find the Anopheles\_Metadata\_Manifest\_V4.0 [HERE](https://docs.google.com/spreadsheets/d/16x9xOhQBKz8xUyfwnRwP_8gmS58nNJFozEsrB6UWu_c/edit?usp=sharing)

## This links to a “view only” copy of the manifest. Please do not request edit access as it will not be granted. You should make a copy of the Google Sheet (File > Make a copy) and work there online or offline (File > Make available offline), or download it as an Excel file (File > Download > Microsoft Excel (.xlsx)). It is important to save the manifest with its original formatting (i.e., not as a .txt or .csv file) otherwise you will lose the drop down menu terms. You may have some issues if you are working on the Google Sheet directly in a browser other than Chrome.

## 

## OVERVIEW

This document contains column by column instructions for capturing metadata at a single sample level (a sample can be a whole mosquito in a 96-well microtiter plate well, a part of a mosquito in a well, or DNA from a mosquito). To accomplish this, the manifest contains two data entry tabs:

TAB 1: “Contributors” records the information for all people who contributed to the batch of samples recorded in the manifest, for example, the collection, identification, and/or preservation of samples.

TAB 2: “Metadata Entry” records information on each individual sample/specimen.

There are also two further tabs:

TAB 3: “EXAMPLE Metadata Entry” shows some accepted examples of data entry to help show what is permitted. This example tab has three plates entered – the first plate (DACH\_001, all 96 entries highlighted in blue) has extensive metadata and represents an ideal case when every sample has complete information. The precise GPS coordinates for each individual were not captured, as a whole village was sampled, so “SAMPLING\_LOCATION\_SIZE” reflects this imprecision by noting 1km2. The second plate (DACH\_002, all 96 entries highlighted in green) has just the mandatory metadata entry required – none of the purple or white fields have information added. This is the minimal information required to process samples and where possible, we recommend adding entries for purple and white headed columns. The third plate (DACH\_003, all 96 entries highlighted in purple) is a situation in which every sample has extensive metadata but it is the same for every specimen because these were collected on the same day from the same location.

TAB 4: “Well IDs” contains for convenience the well numbers on a plate (when plating down columns) so this does not need to be manually entered but can be copy/pasted into TAB 2 field “TUBE\_OR\_WELL\_ID” for each submitted plate. Please note that specimens must be plated down columns, not across rows. If you elect to plate across rows (we do not provide 12-strip caps), you must ensure the specimen is linked to the correct well ID in the manifest (e.g. A1, A2, A3… rather than A1, B1, C1..).

TAB 5: “Data Validation” is solely for populating drop down menus and can not be modified without seeking confirmation from anospp@sanger.ac.uk. It is possible to see all fixed term options in advance of adding metadata by viewing this tab. This tab has been locked for editing on Google Sheets, however when this is downloaded as an Excel file, it is no longer locked. If using in Excel, be careful not to change any terms in this tab otherwise it will affect the drop down menus in TAB 2

**Orange** metadata fields (i.e. columns) are mandatory – we must have meaningful information to proceed with sample processing. **Purple** columnsare strongly requested, but they are not mandatory, therefore if the information was not collected, the entry can be left empty. For all other fields, we encourage data entry if the data have been collected, but we do not require it. After sequencing, all raw sequence data will be submitted to an open access repository linked to its SPECIMEN\_ID.

## A note regarding contributions

All people who hold primary responsibility for your samples and anyone who contributes to individual collections at a level warranting authorship to resulting publications should have their names added to “TAB 1 Contributors” in the manifest that accompanies each submission of specimens in plates. These fields are also highlighted in **orange** indicating that they are mandatory.

## *Manifest submission guidelines*

To fill in the manifest, you can either make yourself a copy of the Google Sheet (File > Make a copy) and work there online or offline (File > Make available offline), or download it as an Excel file (File > Download > Microsoft Excel (.xlsx)). It is important to save the manifest with its original formatting (i.e., not as a .txt or .csv file) otherwise you will lose the drop down menu terms. You may have some issues if you are working on the Google Sheet directly in a browser other than Chrome.

Note that if you copy/paste, you will overwrite any data validation and the error flags that pop up to let you know you have entered data incorrectly will no longer occur, so please be careful when you use copy/paste to enter metadata to ensure that your entries meet the requirements (e.g. “*FEMALE”*, not “*F”* or “*fem”*). We recommend first doing a manual entry of several specimens so that you can take advantage of the data validation and identify where there are likely to be issues when using copy/paste.

Before we are able to accept any samples, your completed manifest must go through a validation process. When you have completed filling in your manifest, but **prior to shipping your samples,** please send the manifest (as a link to a Google Sheet or as an Excel file) to [anospp@sanger.ac.uk](mailto:anospp@sanger.ac.uk) to go through this validation process at Sanger. This ensures that all fields are complete and interpretable for every submitted sample. If any issues with the information provided are identified (e.g., missing mandatory entries, duplicate rows, incorrect date formats) the sample manifest will be returned to you along with a list of detected issues and you can resubmit once you have resolved these issues.

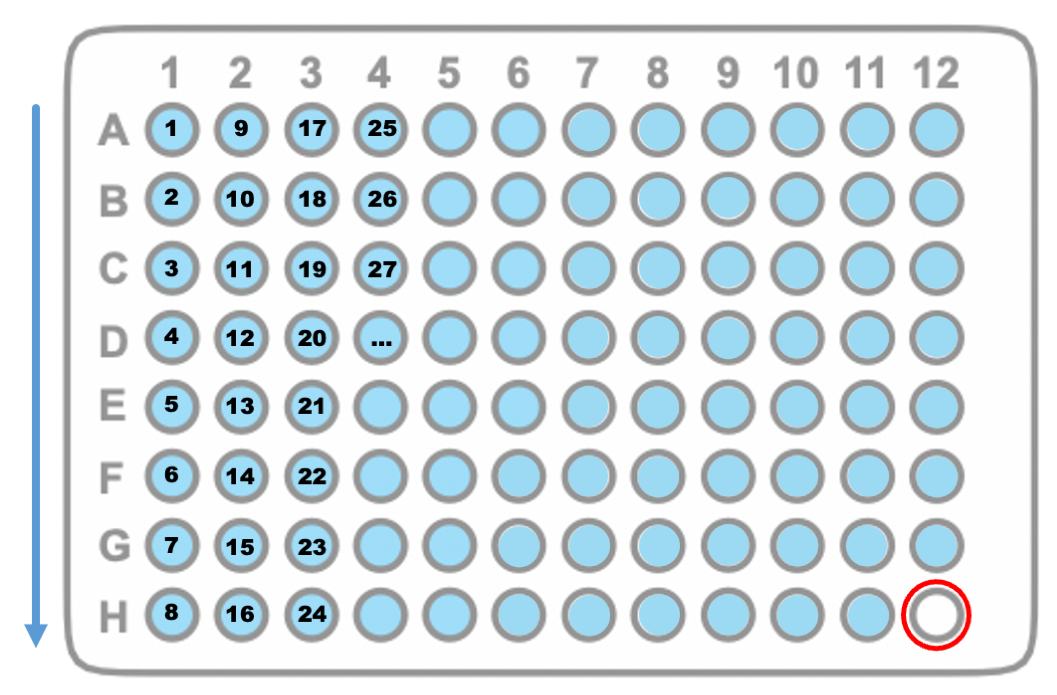
Please carefully read the guidance in this SOP for each field, and attempt to get your submitted manifests as close to the guidance as possible to avoid many iterations of validation.

## A note regarding sample submission

### We can only accept samples in 96-well microtiter plates. Prior to sample collection, please contact us to arrange the shipment of strip caps (Fisherbrand Domed 8-Cap Strips P/N 14-230-231) and pre-labelled plates (ThermoFisher Scientific P/N AB2800) if you cannot supply them yourself. Both DNA and specimens should be shipped in these plates with strip caps properly sealed with a strip cap tool to prevent evaporation. If needed we can also provide a capping and decapping tool together with the plate shipment.

### Plating specimens/ samples

1. Before beginning, ensure your plate is labelled according to the instructions you have received from us and that it is in the correct orientation such that the first well is clearly A1 and well H12 is in the bottom right. Plate insects in a vertical direction in 96-well plates working down columns as pictured in the schematic below (e.g A1-H1 followed by A2-H2 and so on) until all blue wells are filled. We do not typically accept partially filled plates so please ensure that plates are filled entirely with 95 insect specimens and only position H12 is left empty. If there is a need to submit a partially filled plate, please contact us [anospp@sanger.ac.uk](mailto:anospp.info@sanger.ac.uk) to discuss further. Plates should be sealed with tightly fitting 8-strip caps. Films are not recommended unless you have a heat sealer as they tend to peel off and evaporate/leak.



1. Every submitted plate should contain exactly 95 samples. **Please always leave the final well (H12) empty of specimens –** this well must contain the PRESERVATIVE\_SOLUTION but no biological material. If your plates have been pre-loaded and delivered to you by us, there may be a dye or a mark on the well to remind you to keep it empty. If your plates are provided by you, we strongly recommend taking a sharpie and circling the well H12 to ensure correct orientation prior to adding insect specimens. The only fields that must have entries for the empty well H12 are RACK\_OR\_PLATE\_ID, TUBE\_OR\_WELL\_ID, PRESERVATIVE\_SOLUTION (which should be the same as the rest of plate), and ORGANISM\_PART, which must be entered as “*NOT\_APPLICABLE*” as this will be how we identify empty wells in later data analysis.

### Metadata entry

1. The manifest has drop-down menus in some fields. If you discover a missing term in the drop-down menus, new terms can be suggested by contacting us. Please only request new terms if the absence of the term is likely to affect many samples.
2. **Information must be entered for all fields below with orange** **highlighted fields** (column headers)**.** These fields are represented by cells with orange fill. The fill will go white when an entry has been made to help you identify which mandatory fields still require data. All fields with orange highlighted namesare mandatory and must be filled with the appropriate information. For all other fields, if you have collected the information, please add it. If not, you may leave these fields empty. It is permitted to add information for a subset of specimens on a plate.
3. **All dates in the manifest must be formatted consistently as YYYY-MM-DD** (ISO 8601). If only year or year/month are known, use **YYYY** or **YYYY-MM**, respectively.
4. In fields that are “free text” we ask that you use only the core alphanumeric characters, plus full stop “.”, hyphen “-”, underscore “\_” and spaces (summarised in coding parlance as ” **-\_.a-zA-Z0-9**”). Please avoid “|” (the vertical pipe symbol) except where we indicate it should be used to separate elements in a list. Please **do not** use “special characters” (such as other punctuation and “logical” marks: “#”‘;:?!@\*()[]{}/\,=+”, etc.).

**When you have completed the metadata entry and before you send any samples, please email your manifest or a link to it to** [**anospp@sanger.ac.uk**](mailto:anospp@sanger.ac.uk). Please name your manifest PARTNERCODE-YYMM e.g. DACH-2105; PARTNERCODE is assigned during partner onboarding, and YYMM is the year and month of your initial manifest submission to the email address above rather than the timing of the collection or shipment.

# 

# Tab1: Contributors

All contributors to the sample collection and the overall study should have their names and affiliations listed in this tab. Please list all people who contributed at a level warranting recognition to samples listed on TAB2. This may include, for example, people who are responsible for the collection, identification, preservation of samples or have overall responsibility in the project.

1. **SURNAME:** First letter capitalised, rest lower case (e.g., Darwin)
2. **FIRST\_NAME:** First letter capitalised, rest lower case (e.g., Charles). If middle initials should be included, please add these here (e.g. Charles R. ).
3. **PRIMARY\_AFFILIATION:** Add the primary affiliation for each individual. For those without an official affiliation (e.g. volunteers), partners should discuss the use of their affiliation as the work was done on their behalf. Additional affiliations and complete addresses will be requested at the time of publication by emailing everyone on the Contributors Tab.
4. **EMAIL\_ADDRESS:** Add a reliable email address as this is the primary route by which we will contact you for authorship queries.
5. **CONTRIBUTION:** List the contribution(s) made by the individual (e.g., collected specimens). Please list “Primary Contact” as an entry here for the person responsible for answering queries on the specimens.
6. **CONFIRMATION:** For every shipment, the primary investigator leading the project must confirm that all samples contained within the shipment have been collected with local permissions and under local ethical guidelines. Please write “YES” in this column to confirm that these standards have been met and that the necessary regulatory compliance documents have been obtained and are available to you. We are not able to accept samples that do not meet these standards. These may include landowner permission, veterinary pathogen sampling permissions, and/or Nagoya compliance. This is an important check that ensures that permissions were granted to collect and transfer the specimen for this research purpose. The sample provider should ensure this documentation is obtained, and that copies of the relevant paperwork are shared with Sanger where necessary and as stipulated, for example, by regulations/approvals or licensing authorities.

# Tab2: Samples

# Column by column instructions for the Metadata Entry tab

* 1. **SERIES:** This field simply is a series of numbers that should reflect the total number of wells submitted, e.g., if 24 plates are submitted, each with 95 samples, then the series will begin at 1 and end at 2304 (the control well/blank sample for each plate must also have a row in the manifest). Partners are expected to ship a minimum of 8 plates of samples for processing, and where possible, in units of at least 8 plates (e.g. 8, 16, 24 plates) because we process 8 plates at a time for a single MiSeq run. The sheet contains enough rows to support a shipment batch of 50 plates (e.g. 4750 mosquitoes) and can be extended if a shipment batch is larger than this.
  2. **RACK\_OR\_PLATE\_ID:** Each plate must be labelled with a sticker that has a human readable label and a scannable barcode. The human readable label will contain your Partner Code, which we will assign to you, and a numeric plate identifier (e.g. DACH\_001 would represent **CH**arles **DA**rwin’s first plate). We will provide you with the stickers to label your own plates or with labelled plates if you require these materials from us.
  3. **TUBE\_OR\_WELL\_ID:** This field should have a record for each well in a plate. Adding the WELL\_ID for samples in plates can be expedited by copy/pasting from the appropriate column in TAB 3, the Well IDs tab. The first 8 plates are completed for you as we typically expect 8 plates at a minimum for a shipment batch. Please ensure that you plate your specimens going down columns as in the schematic above and not across rows.
  4. **PRESERVATIVE\_SOLUTION:** Suspension liquid used to preserve each sample in the 96 well plate, which should typically be “*100%\_ETHANOL*”. This field has a controlled vocabulary: select the correct option from the drop down menu for the first entry. To fill multiple samples with the same information, click on the cell containing the data entry and then click on the little blue square that will appear at the lower right corner of the cell (in Google Sheets) and drag the square down to fill any remaining cells. If a different percentage of ethanol has been used, select “*V%\_ETHANOL*” and add the percentage (e.g. 85% ethanol) to the field PRESERVATION\_APPROACH. If none of the drop-down terms are suitable, or if DNA is provided, select “*OTHER*” and add this information to PRESERVATION\_APPROACH.
  5. **ORGANISM\_PART:** A description of the exact tissue(s) in the tube or well. This field has a controlled vocabulary: use the drop-down menu. For the empty well required on each plate that should contain only the preservative solution (or for any other wells that do not contain specimens), select “*NOT\_APPLICABLE*” (H12). Entering “NOT\_APPLICABLE” for empty wells is critical as it informs us that there is no sample expected in the well.
     + Note that if partial specimens are submitted, you should choose the appropriate combination of tissues from the drop-down menu (e.g., “*HEAD | THORAX*”). This will result in a data validation error (i.e. a little red triangle appearing at the top right corner of the cell), but as long as the parts are spelled as listed, this is permitted. If some damage has been done but the specimen is largely complete (e.g., missing some legs or wings), then select “*WHOLE\_ORGANISM*”.
     + If the sample is shipped as a DNA extract, it is still important to select the tissue from which the DNA was extracted here. Then, further information in the three dedicated DNA columns at the end of the manifest will be required. Note that any shipment of DNA should be discussed in advance as insect tissue is typically expected.
  6. **DATE\_OF\_COLLECTION:** The date of the sample collection, with year, month and day specified in this order using hyphens (YYYY-MM-DD). It is important to provide the complete date, but if only the year or the year and month are known, then enter YYYY or YYYY-MM, respectively. If the collection spans multiple days, we request that the **end** date is entered here rather than the start date. If your collection ends precisely at midnight, we consider midnight to be the first minute of the day (00:00 as opposed to 24:00) and thus the end date should reflect this.
  7. **COLLECTION\_COUNTRY:** This must use the accepted country name, which can be looked up here <https://www.insdc.org/country.html> .
  8. **COLLECTION\_LOCATION:** Where possible, this should use your country’s administrative districts to as fine of a resolution as you can provide ranging from least to most specific and separated by | character, e.g. “*England | East of England | Cambridgeshire | Hinxton | Wellcome Genome Campus | East Pond*”. It is important to give the name of the specific collection site whenever possible. If the specimen is from a laboratory colony, give the current location (e.g. Imperial College London) rather than information on the original collection, but indicate original collection location in **OTHER\_INFORMATION**.
  9. **DECIMAL\_LATITUDE:** In decimal degrees, between -90 and 90. Ideally, locations should be specified to 3 decimal places (<https://en.wikipedia.org/wiki/Decimal_degrees>) giving 111 m resolution to the location from which the sample was collected. However, the field “SAMPLING\_LOCATION\_SIZE” (Column K) captures the accuracy of the resolution provided in square kilometres. If GPS coordinates were not taken at the time of collection, then please add the GPS coordinates for the region afterwards. You can do this by dropping a pin in Google Maps on the location where the specimen was collected and clicking on the pin to reveal coordinates.
  10. **DECIMAL\_LONGITUDE:** guidance same as latitude, except from -180 to 180.
  11. **SAMPLING\_LOCATION\_SIZE:** This field captures the accuracy of the collection location GPS coordinates you provided in square kilometres. It is a drop down menu – please round up to the nearest accurate area represented by the GPS coordinates you provided (1m2, 10m2, 100m2, 1km2, 10km2, 100km2).
  12. **PREDICTED\_SCIENTIFIC\_NAME:** The latin binomial/combined genus and species name with a space in between. If the specimen has not been identified at least to species group level using morphology or molecular methods, it is permitted to leave this field empty as long as plates are filled with 95 specimens and well H12 is left empty. If the well does not contain a sample (e.g. H12), again this field can be left empty as long as ORGANISM\_PART has been registered as “NOT\_APPLICABLE”.
  13. **SPECIMEN\_IDENTITY\_RISK:** Y/N field to indicate if there is any risk that the specimen is part of a species complex or group where it can be difficult to be certain of species identity. Enter “*N*” (i.e. no risk) if there is high confidence that the specimen has been identified to the correct species level as recorded in PREDICTED\_SCIENTIFIC\_NAME. If there is high confidence that the species listed is correct, please provide information on your species identification methods in the IDENTIFIED\_HOW column. If you can identify the specimen to a species group or complex, but perhaps not all the way to species level, enter “*Y*” (i.e., yes there is a risk the specimen is not accurately identified to species level resolution).
  14. **IDENTIFIED\_HOW:** If you have added an entry to PREDICTED\_SCIENTIFIC\_NAME, please indicate what method(s) were used to identify the specimen to the nominal species (e.g., morphology). This is free text and can include reference to an authoritative key if possible.
  15. **LIFESTAGE:** The lifestage of the specimen from which the sample was derived at the time it was preserved (if this is different than the lifestage when collected e.g. if larvae were collected in the field and reared to adults in the lab, the entry here should be adult). This field has a controlled vocabulary, please use the drop-down menu. If you would like to provide more detailed lifestage information than is available in the drop-down menu, add this to the OTHER\_INFORMATION column (e.g., “*L3 stage larva*”).
  16. **SEX:** The sex of the specimen from which the sample was derived. This field has a controlled vocabulary: use the drop-down menu. If the sex of the organism is not known, you can leave this field empty.
  17. **TIME\_OF\_COLLECTION:** Time of day at which the collection **ended** in local time in 24-hour clock format, with hours and minutes separated by colon e.g. “*13:35*”, “*04:53*”, etc. We do not accept 24:00 for midnight and you must use “*00:00*” (first minute of the day) if your collection ends precisely at midnight. This should be in local time. If your collection spans midnight -- for example, if it began at 23:30 on April 1, 2021, and then ended at 00:30 on April 2, 2021, your entry for COLLECTION\_DATE should be “*2021-04-02*”, for TIME\_OF\_COLLECTION should be “*00:30*”, and for DURATION\_OF\_COLLECTION (below) should read “*PT1H*”. For a trap when the precise collection time of each specimen is unknown, record the time at which you end the catch and in the next field, record the duration of the catch (e.g. “*PT24H*” for a 24 hour catch). If the time was not recorded, you can leave this field empty.
  18. **DURATION\_OF\_COLLECTION:** This field captures the duration of the catch using ISO 8601 standards. The format is complicated and described below in detail. Briefly, if the collection time you provided in the preceding field (TIME\_OF\_COLLECTION) is accurate to the minute level, enter PT1M (e.g. you caught the mosquito at exactly that time). If it is accurate to the hour, enter PT1H (e.g. you caught the mosquito some time in the hour preceding the time recorded above), and if it is accurate to a 10 hour window (e.g. if you caught the mosquito some time in the 10 hours preceding your entry for TIME\_OF\_COLLECTION), enter PT10H, etc.

More detailed instructions: The format is “*P[n]Y[n]M[n]DT[n]H[n]M[n]S*”. In this representation, the [n] is replaced by the value for each of the date and time elements that follow the [n]. The capital letters *P*, *Y*, *M*, *W*, *D*, *T*, *H*, *M*, and *S* are designators for each of the date and time elements and are not replaced.

*P* is the duration designator (for *period*) and is always placed at the start of the duration representation.

* + *Y* is the year designator that follows the value for the number of years.
  + *M* is the month designator that follows the value for the number of months.
  + *W* is the week designator that follows the value for the number of weeks.
  + *D* is the day designator that follows the value for the number of days.

*T* is the time designator that precedes the time components of the representation.

* + *H* is the hour designator that follows the value for the number of hours.
  + *M* is the minute designator that follows the value for the number of minutes.
  + *S* is the second designator that follows the value for the number of seconds.

For example, if a trap or collection period was running for 23.5 hours, the record here would be “*PT23H30M*”, or if it was running for 11 days and 2 hours, this would be “*PT1W4DT2H*”. For a 30 minute human landing catch interval, this would be recorded as “*PT30M*”. If an exact catch time is recorded in TIME\_OF\_COLLECTION, then DURATION\_OF\_COLLECTION should have “*PT1M*” entered to indicate this was the exact minute of collection. Note that “T” is critical as “*P1M*” would be read as 1 month duration. If the duration was not recorded, you can leave this field empty.

* 1. **COLLECTION\_METHOD:** This field has a controlled vocabulary; pick the best fitting term from the drop down menu. If none of the available terms describe your collection method, select ”*OTHER*” and describe the method in the column DESCRIPTION\_OF\_COLLECTION\_METHOD or request an additional term to be added.
  2. **OUTDOORS\_INDOORS:** For any specimen collected inside a building or home, please select “*INDOORS*”, otherwise select “*OUTDOORS*”. If this was not recorded, you can leave this field empty.
  3. **DESCRIPTION\_OF\_COLLECTION\_METHOD:** A detailed as possible description of the collection methods, e.g. “*caught with fibre net within densely wooded area, and immediately placed into the collection container”* or “*manually collected whilst feeding on a cow inside a shed and immediately placed into the collection container”*. If the specimen was collected as larvae and reared to adult, mention this here.
  4. **PRESERVATION\_APPROACH:** This is free text but should summarise everything that has happened from the point of collection to the point of sample being placed in the well in the following format: “*kill method; preservation 1: method, duration, temperature, preservation 2: method, duration, temperature*”. For example, if a specimen was killed in ethanol, held in a tube with desiccant for 13 years at RT, then moved to ethanol at RT for shipment, the entry would read: “*kill method = 100% ethanol submersion for 3 minutes; preservation 1 = placed in eppendorf with silica gel and cotton wool for 13 years at RT; preservation 2 = moved to 100ul 100% ethanol at RT on 2021-03-01*”. A single preservation approach is fine (e.g., *“kill method = 100% ethanol submersion for 3 minutes; preservation = transferred to a plate containing 100 ul of 100% ethanol per well and held at RT”*). Please include volumes and percentages where relevant.

**Terms below here are fully optional and can be left empty if there are no entries**

* 1. **BLOOD\_MEAL:** Indicate “*Y*” (for YES) if the collected specimen had a visible blood-meal or was known to have had a blood meal. “*Y*” should be selected even if the abdomen has been excluded from the well to be used for other purposes and there is unlikely to be host or parasite DNA present. Otherwise, put “*N*”. For blood-fed samples where the feeding source is known, enter the scientific name of the feeding source in the OTHER\_ORGANISMS field.
  2. **GRAVIDITY**: Indicates whether or not (Y or N) visible eggs are observed in the insect’s abdomen. If gravidity has been assessed, please add the gravidity status for the sample even if the eggs were not included in the submitted specimen.
  3. **HABITAT**: Any comments about the location, habitat or substrate, *e.g. “damp mossy ground in moderate shade*“ or “*indoors in air conditioned office space*“. Also of interest is whether there are livestock nearby and what kind.
  4. **DATE\_OF\_PRESERVATION**: Some mosquitoes may be kept alive for a period of time for example to study if they were to become infected with *Plasmodium* after a bloodmeal. This field aims to capture whether any days have passed between collection and preservation (i.e., death). If there is more than one preservation this should be the date of the first preservation event when the mosquito was killed. Please use the YYYY-MM-DD format. It is important to provide the complete date, but if only the year or the year + month are known, then enter YYYY or YYYY-MM, respectively. If the date of preservation (i.e., death) is the same as the date of collection, this field may be left blank.
  5. **ELEVATION:** Altitude above sea level, supplied in metres. Do not supply the unit, e.g. use “*200*” for 200 m above sea level, “*100-200*” for 100-200 m range above sea level, etc. Please supply elevation of water surface for inland water bodies.
  6. **WHAT\_3\_WORDS:** Information to geolocate the sample area to a 3 m square, mapped by what3words.com. This can be used to give more precise resolution to distinguish between nearby sampling locations, e.g. different puddles for larvae collection or mosquitoes caught in different houses. Please include the three backslashes as part of the entry, e.g., “*///protected.cheetahs.slippery*”.
  7. **OTHER\_ORGANISMS**: Scientific names of other species expected to be present in this specific sample, taking consideration of the ORGANISM\_PART. If multiple species are expected, separate their names using “|” (the vertical pipe symbol). If the sample is collected whilst feeding, then this term should be used to record the scientific name of the feeding source (for instance the name of the animal a mosquito was feeding on). Only enter this information if this specific sample was observed to be interacting with the OTHER\_ORGANISMS.
  8. **BIOASSAYS:** If the specimen went through an insecticide resistance bioassay of any sort, please provide more details here, e.g. standard WHO/CDC bioassays, paired with a synergist or intensity bioassays. This is a free-text field. Example entries might be “The WHO 1998 bioassay procedures were used to test DDT at 4% concentration for 60 minutes. 100 mosquitoes were tested and the percentage mortality was 89%. This mosquito survived”. or “150 mosquitoes were tested with permethrin and resulted in 75% mortality. The CDC bottle bioassay was used at 21.5 μg/bottle concentration with exposure of 30 minutes”. If there were no bioassays completed on the specimen, please leave this field empty.
  9. **COLLECTOR\_SAMPLE\_ID**: This is the unique name assigned to the sample by the COLLECTOR. This is a free text field, but please do not use spaces or special characters, other than hyphens, underscores, and full stops (i.e., do not use “#”, “!”, “^”, “\*”, etc.)
  10. **OTHER\_INFORMATION**: Free text field for further relevant information not captured by the other fields. If the samples were reared in a laboratory, either from an existing colony or strain or as offspring from wild-caught organisms, please note this here. Occasionally, the relationships between submitted samples may be known and important. This can be reflected in the COLLECTOR\_SAMPLE\_ID and described here (e.g. “*AB43\_F1*” and “*AB43\_Mother*”).
  11. **MISC\_METADATA**: Please discuss this with us ([anospp@sanger.ac.uk](mailto:anospp@sanger.ac.uk)) before filling in the manifest if additional data for your samples exist but are not captured by any existing fields. In such a case, we should make every effort to collect and standardise additional metadata. This column can be copy/pasted to create as many additional columns as required to capture any additional metadata of interest, for example, this might include temperature, humidity, windspeed, karyotype, urban vs sylvatic, parous vs nulliparous, virgin vs mated, or bioassay outcome. Please change the field from MISC\_METADATA to the proposed name of the new field, and standardise your entries using ALL CAPs and no spaces.

The final metadata fields described below are only relevant if you are submitting DNA extracts.

* 1. **DNA\_EXTRACTION\_DESCRIPTION**: Free text field for a brief description of the technique or kit used to extract and purify DNA from the sample (e.g. DNeasy, NextTec) and the storage conditions of the DNA since extraction. If non-proprietary extraction buffers or procedures have been used, a brief but informative description should be included listing the buffer components and reagents, e.g. “*samples incubated at 56C overnight in 60uL of lysis buffer (200mM Tris, 25mM EDTA, 0.4mg/ml Proteinase K, 0.05% Tween-20), DNA purified using the MinElute 96 UF PCR Purification Kit following the manufacturer’s protocol with modifications (included water wash step, eluted in 40ul). DNA stored at -20C since 2010*”. Indicate what the DNA has been eluted in (e.g. TE, water) and add the same liquid into well H12 (the “blank sample”). In cases where samples have been put directly into lysis buffer, but no DNA quantification has occurred and no volume has been removed for other purposes, leave this field empty.
  2. **DNA\_EXTRACT\_VOLUME\_PROVIDED**: Free text field to input the provided sample volume in microlitres (µL). The volume should range between 10 µL (minimum) to 100 µL (maximum), e.g. if 20 µL are provided, add “20” here.
  3. **DNA\_EXTRACT\_CONCENTRATION**: Free text field to input the DNA extract concentration estimates in ng/µL, e.g. if the concentration is 10 ng/µL, add “10” here.

# The Post-validation Process

All fields listed above require metadata to be filled in by specimen providers. Once validation has been completed, additional fields are created as described below and a copy of this final manifest will be returned to you.

**TAXON\_ID:** This is recorded in all cases as “*unidentified*” 32644 or “*blank sample*” 2582415.

**SCIENTIFIC\_NAME:** This is recorded in all cases as “*unidentified*” if there was a specimen in the well. This is a formality to ensure that we are able to update to the correct name after sequencing is complete. Empty wells will then be registered as “*blank sample*” to differentiate from “*unidentified*” in the validation process.

**SPECIMEN\_ID:** This is a unique identifier given to each specimen through combining RACK\_OR\_PLATE\_ID + ‘\_’ + TUBE\_OR\_WELL\_ID (e.g., DACH\_001\_A1)

**PREDICTED\_TAXON\_ID:** This is the NCBI Taxonomy identifier assigned to the specimen based on the PREDICTED\_SCIENTIFIC\_NAME – if no scientific name is provided, this will be recorded as 32644. If a species name is provided, the correct TAXON\_ID will be added to this field.

**GAL, SYMBIONT, REGULATORY\_COMPLIANCE, and HAZARD\_GROUP** may all also be added upon completion of validation because they are required by our internal sample tracking system.