



Guidelines for submitting MCZ Catalog Numbers and to supplement Specimen Data for GenBank

When submitting data to GenBank through tools such as Sequin or BankIt, MCZ Catalog numbers must be included. Secondary identifiers associated with the sequence may also be documented if desired. MCZ Catalog numbers and other identifiers are recorded on the **Organism Subpage**, where all supplementary information regarding the source organism is documented. Note that multiple modifiers can be selected for each sequence submitted.

Record the MCZ Catalog Number

- Select the “**specimen-voucher**” modifier from the pull-down list. This modifier describes the *primary* identifier of the physical specimen from which the sequence was obtained, and is intended for use in cases where the voucher/sample is still available (e.g., museums, herbaria, frozen tissue collections). Selecting the specimen-voucher modifier will generate three separate fields so that the MCZ Catalog number information can be entered in the following required format:

institution_code:collection_code:specimen_id (e.g., MCZ:Mamm:12345). In order to properly link the GenBank sequence to the corresponding specimen record in MCZbase, all MCZ Catalog numbers should be complete and include any prefixes and original formatting (e.g., MCZ:Herp:R-2244).

Below is the list of relevant Collection Codes for the different MCZ Collections. Note that the first letter of each collection must be capitalized:

| | |
|----------------------|------|
| Cryogenic | Cryo |
| Entomology | Ent |
| Ichthyology | Ich |
| Herpetology | Herp |
| Invertebrate Zoology | IZ |
| Malacology | Mala |
| Mammalogy | Mamm |
| Ornithology | Orn |

GenBank records for material submitted to GenBank using the proper MCZ Catalog number format as noted above, will be automatically hot-linked to the proper MCZbase specimen record (and the MCZbase record will be hot-linked back).

Record Other Identifiers

As suggested by GenBank, “bio-material” and “isolate” modifiers can be used to capture additional numbers associated with a sequence or voucher specimen that are secondary to the MCZ Catalog number.

- Secondary identifiers such as DNA Reference Codes or numbers specifically associated with tissues or blood samples, such as those recorded on vials, can be entered on the Organism Subpage using the modifier “**bio-material.**”



Although three fields will display when the “bio-material” modifier is selected, only record the secondary ID (including any prefixes or alpha characters) in the “specimen_id” field (do NOT enter institution_code or collection_code in this case).

- Secondary identifiers associated with the voucher specimen such as field numbers, collector numbers, and original numbers should be recorded in the “**isolate**” field.

Again, only enter the specimen secondary ID in the “specimen_id” field (do NOT enter institution_code or collection_code).

Additional Data

- It is suggested that information regarding type of tissue used to obtain the sequence data (e.g., heart, muscle, liver, ground whole organism) be recorded as a source modifier using the “**tissue-type**” field.
- Additional comments and remarks can be recorded using the “**notes**” modifier.
- Refer to the relevant GenBank submission tool documentation for further instructions and a glossary of all available modifiers.

Note that by submitting the MCZ Catalog number in the “**specimen-voucher**” field, the GenBank sequence record will link directly to the corresponding MCZbase specimen record, where the voucher specimen’s collecting data and biological attributes can be readily accessed. It is up to the researcher’s discretion as to whether or not to include such informational redundancies when submitting locality and biological data to GenBank on the Organism Subpage.