



Providing Source Information in your Submission

Created: April 6, 2011; Updated: November 3, 2014.

Source information for Samples Collected in the Field

What is an Isolation_Source?

I have isolated sequences from samples I obtained in the field, and have been asked to provide an `isolation_source` — what is an `isolation_source`?

The `isolation_source` is a modifier that describes the physical, environmental and/or local geographical source of the biological sample from which a sequence was derived (e.g. soil, sediment, ocean water, lake water, forest debris, soil from outside a specific chemical factory, gasoline polluted soil, etc.)

Country and Latitude/Longitude Information

How specific should my latitude longitude (`lat_lon`) information be?

Provide the `lat_lon` in decimal degrees that include compass direction (e.g. 39.7 N 42.1 W). Also provide the `country of origin` for your sample if you have it.

Since I don't know the latitude and longitude (`lat_lon`) of my samples, should I give the latitude/longitude based on the `lat_lon` of a near-by city or landmark?

- **Do not provide `lat_lon` information** if you have to determine the latitude and longitude of the sample site (based on a nearby city or landmark) after the sample was collected.

If you do not have the `lat_lon` information for your sample, provide the `country of origin` for your sample. Include the specific locality where the sample was obtained (if known), using the following format:

```
/country="country:sub_region"
```

For example:

```
/country="Canada:Vancouver"
```

Or

```
/country="USA: Bethesda, MD"
```

- **Provide `lat_lon` information only** if you recorded it at the time you collected the sample. Also provide the `country of origin` for your sample if you have it.

Additional Collection Information

How do I add (extensive/detailed/complicated) isolation, location, or other information to my submissions for the organisms from which I isolated my sequences?

Detailed or extensive information describing the location where your organisms were isolated or detailed organism descriptions can be included in the following source modifiers:

- /authority

The /authority modifier should include the complete name of the organism, followed by the authority information.

For example:

/authority=Elekmania picardae (Krug & Urb.) B.

/authority=Avena sativa L.

- /collected_by

The /collected_by modifier should report the name(s) of the specific person(s) who collected the organism from which the sequence was obtained.

For example:

/collected_by=Fred MacMurray

/collected_by=A. Hitchcock and F. Zeferelli

- /collection_date

The /collection_date modifier must be in the format DD-Mon-YYYY or Mon-YYYY or YYYY.

For example:

/collection_date=30-Sep-2008

/collection_date=Sep-2008

/collection_date=2008

- /country

The /country modifier can also include province, state, region, oceans, or other locality names. The name of the country (or ocean) must be provided first, followed by a colon (:) before the additional location information.

For example:

/country=USA: Lancaster County, PA

/country=Canada: SW coast of Newfoundland

/country=USA: Syracuse State Park in upstate New York

/country=Atlantic Ocean: 24.5 miles east of Bermuda

/country=Pacific Ocean: Stubing Marine Station

You can find a [list of INSDC approved country names](#) online at NCBI.

- /identified_by

The /identified_by modifier reports the name(s) of the specific person(s) who identified the TAXONOMY of the organism from which the sequence was obtained. This does not mean the

person(s) in the laboratory who identified the submitted sample.

For example:

/identified_by=James Cagney

/identified_by=K. Hepburn and S. Poitier

- /isolation_source

The /isolation_source modifier describes the physical, environmental, and/or local geographic location where the organism was isolated.

For example:

/isolation_source=cow rumen

/isolation_source=abandoned silver mine 20 miles NW of Las Vegas

/isolation_source=roadkill on Old Sulphur Mill Road

/isolation_source=activated sludge from bioreactor

/isolation_source=runoff from chicken farm

- /lat-lon

The /lat-lon modifier should be in a decimal degree format using single letters that denote direction and should report the latitude and longitude measured at the site and time of collection.

For example:

/lat_lon=47.68 N 33.75 W

/lat_lon=28.82 S 12.50 E

Note:

- The latitude and longitude **cannot be estimated** from a map or GPS device **after** the collection is complete.
- The specific country (or ocean) should also be reported as a/country modifier.

Additional organism metadata that do not fit into any of the available [modifiers](#) may be appropriate in a [Structured Comment](#).

Eukaryotic Source Material

What kind of source information do I provide with a Eukaryotic Set submission?

- **If the source material originated in a museum or other reference collection, provide the following information only** if the sequence you are submitting was obtained from a sample that you retrieved directly from the indicated museum/collection, or the sequence was obtained from a sample that you deposited in the indicated museum/collection:
 - The specimen voucher number for each different source material used. Specimen vouchers provide a means to verify the identity of a taxon and are a source for additional molecular analyses.
 - **See the “Museum/Reference Collection Source Material” section** to see the types of information accepted for museum/reference collection source materials).
- **If your source material does not come from a museum or other reference collection, and has no specimen voucher number**, provide any of the following information about your source material(s):
 - Cultivar, strain, isolate, or breed
 - Germplasm, seed, or stock center accession number (use biomaterial modifier)

- Collection number, locality, and/or date

Rat/Mouse Source Material

What kind of source information do I provide with a set of mouse/rat sequences?

In addition to the information for eukaryotic source material(s), provide the mouse/rat strain name for each mouse or rat strain submitted. If you do not know the strain name, please tell us at the time of your submission.

Bacterial and Archaeal Source Material

What are strain identifiers? Why do I need to provide one for each cultured bacterial/archaeal sequence submission?

Strain identifiers serve to distinguish your culture from other related isolates of the same species obtained in your lab or elsewhere.

Note: your isolates do not need to be deposited in a culture collection in order for you to create strain identifiers for them.

Each cultured bacterial/archaeal sequence you submit should have a unique strain identifier associated with it.

- If your culture comes from a culture collection and has an established identifier, use it as the strain identifier.
- If your culture does not come from a culture collection, you can create **a strain identifier by using anything that identifies the particular culture from which the sequence was obtained:**
 - Any distinguishing identifier you use in the laboratory
 - A string of numbers and/or letters

Since the species name of your bacterial/archaeal isolate alone does not uniquely identify a particular culture, you must provide an identifier for a particular culture of the bacterial/archaeal isolate.

If a species has not yet been assigned to your isolate, you still must provide an identifier for it using the suggestions for creating a strain identifier mentioned above.

How do I provide unique source information for my bacteria/archaea submission?

If you are submitting a number of different sequences isolated from different strains/isolates/clones, provide the information as a spreadsheet or [tab-delimited table](#):

SeqID	strain
Seq01	ABC
Seq02	CBS 235

etc.

If the same sequence was found in separate strains/isolates/clones, create an additional sequence submission for each source type and submit the group of sequences together.

Or you can provide a tab-delimited table giving a single source in the modifier column (e.g. strain) for the sequence, and then list in a note within the table any additional sources where you found the sequence:

SeqID	strain	note
Seq01	ABC	identical sequence found in strains DEF and GHI

Table continued from previous page.

SeqID	strain	note
Seq02	CBS 235	identical sequence found in strains CBS 236 and CBS 237

etc.

If I'm submitting sequence obtained from uncultured bacteria/archaea, what descriptive information do I need to provide about the source?

If you are submitting sequence from an uncultured source, in addition to the information presented in the question/answer unit about [bacterial/archaeal genomic sequence submissions](#), identify the submission source material as uncultured, and provide the following:

Details describing the isolation source (environmental conditions) where the bacteria/archaea was isolated and a unique clone identifier. If you are submitting sequences isolated from multiple conditions and/or locations, provide the information as a spreadsheet or [tab-delimited table](#):

SeqID	isolation source	clone
Seq01	soil	Qx27
Seq02	ocean water	Qy28

etc.

If the same sequence was found in separate isolation sources, create an additional sequence submission for each source type and submit the group of sequences together.

Or you can provide a tab-delimited table giving a single source in the modifier column (e.g. environment) for the sequence, and then list in a note within the table any additional sources where you found the sequence:

SeqID	isolation source	note
Seq11	soil	identical sequence found in pond water and tree bark
Seq12	ocean water	identical sequence found in sand and in samples collected from surface of coastal boulders

etc.

Viral Source Material

What kind of source information do I include in a viral sequence submission?

Provide the following information with your viral sequence submission in a [tab-delimited source table](#):

- Strain or Isolate
- Serotype or Genotype, if appropriate
- Host
- Country
- Collection_date

See an online [example](#) of the annotation of a viral sequence submission.

Museum/Reference Collection Source Material

What kind of source information do I include in submissions whose sequence is extracted from specimens obtained from museums or from other reference collections?

There are three different source modifiers that define the type of information accepted for museum/reference collection source materials:

- culture_collection
- specimen_voucher
- bio_material

The type of information to submit with your sequence depends upon which of these three modifiers best describes the source material from which you extracted your sequence.

Each of these three modifiers is defined below. Select the modifier below that best describes the source material you have. The specific information you need to provide and examples of that information follow the definition of each source type.

1. **/culture_collection:**

- This modifier is used to annotate the following source material types:
 - Live microbial and viral cultures
 - Cell lines that have been deposited in curated culture collections
 - **Provide the following information only if** the sequence you are submitting was obtained from a sample you retrieved directly from the indicated culture collection, or the sequence was obtained from a sample that you deposited in the indicated culture collection:
 - The institution code for the institution where the culture is housed (mandatory).
 - The identifier of the culture from which the nucleic acid sequenced was obtained (mandatory).
 - Optional collection code.

A searchable database of institution and collection codes is currently being developed.

Note: The institution-code (and optional collection-code) must be taken from the INSDC controlled vocabulary (preselected and predefined authorized terms) that denote the institution/collection where the culture is maintained.

- **The format of the /culture_collection information you provide** can be either of the following:

```
/culture_collection=institution-code:specimen_id
/culture_collection=institution-code:collection-
code:specimen_id
```

- **Example of the /culture_collection information:**

```
/culture_collection="ATCC:26370"
```

If you annotate a sequence **with more than one culture_collection modifier**, this indicates that the sequence was obtained from a sample that was deposited (by the submitter or a collaborator) in more than one culture collection.

Microbial cultures in personal or laboratory collections should be annotated using strain modifiers

2. **/specimen_voucher:**

- This modifier is used to annotate the following source material:
 - **A physical specimen that remains after the sequence has been obtained.**
 - Ideally the specimen(s) is/are housed in a curated museum, herbarium, or frozen tissue collection, but it/they can be housed in a personal or laboratory collection as well. If the specimen was destroyed in the process of sequencing, electronic images (e-vouchers) are an adequate substitute for a specimen voucher that identifies physical remains.

- **Provide the following information only if** the sequence you are submitting was obtained from a sample you retrieved directly from the indicated museum/collection, or the sequence was obtained from a sample that you deposited in the indicated museum/collection:
- The unique identifier of the specimen from which the nucleic acid sequenced was obtained. (mandatory)
- The institution code for the institution where the specimen is housed. Omit the institution code if the specimen comes from a personal collection.
- Optional collection code.
A searchable database of institution and collection codes is currently being developed.
- **Note: The institution-code (and optional collection-code)** must be taken from the INSDC controlled vocabulary (preselected and predefined authorized terms) that denote the institution/collection where the specimen resides.
- **The format of the /specimen_voucher information you provide** can be any of the following:

```
/specimen_voucher=specimen_id
/specimen_voucher=institution-code:specimen_id
/specimen_voucher=institution-code:collection-code:specimen_id
```

- **Examples of the /specimen_voucher information:**

```
/specimen_voucher="UAM:Mamm:52179"
/specimen_voucher="AMCC:101706"
/specimen_voucher="USNM:field series 8798"
/specimen_voucher="personal:Dan Janzen:99-SRNP-2003"
/specimen_voucher="99-SRNP-2003"
```

3. /bio_material:

- This modifier is used to annotate source material in biological collections that do not fit into either the /specimen_voucher or the /culture_collection modifier categories:
- Physical specimens from **zoos**
- Physical specimens from **aquaria**
- Physical specimens from **stock centers**
- Physical specimens from **seed banks**
- Physical specimens from **germplasm repositories**
- Physical specimens from **DNA banks**
- **Provide the following information only if** the sequence you are was obtained from a sample you retrieved directly from the indicated collection, or the sequence was obtained from a sample that you deposited in the indicated collection:
- The identifier of the biological material from which the nucleic acid sequenced was obtained (mandatory)
- Optional institution code
- Optional collection code. If you decide to include the collection code, you must also provide the institution code.

A searchable database of institution and collection codes is currently being developed.

Note: The institution-code (and optional collection-code) must be taken from the INSDC controlled vocabulary (preselected and predefined authorized terms) that denote the institution/collection where the specimen resides.

- **The format of the /bio_material information you provide** can be any of the following:

```
/bio_material= =specimen_id
/bio_material= =institution-code:specimen_id
/bio_material= =institution-code:collection-code:specimen_id
```

- **Example:**

```
/bio_material="CGC:CB3912" .
```

How to Describe Unknown Source Material in Your Submission

Can I use the word “unknown” if I don’t know the organism from which a sequence came?

GenBank cannot process your sequence and assign an accession number to it if the source material is simply described as “unknown”.

Different source materials are described below. Read each description carefully; once you have found the description that best describes the source material you have, the information you will need to provide for the source material follows.

A. Were the sequences you want to submit derived from:

1. Pure culture?

(a culture that contains only one microbial species),

If so, provide the taxonomic lineage as far as you have determined it, and include the strain identifier. We will accept any of the following for a cultured organism: (See Box 10)

2. Enrichment culture?

(use of selective culture media to enrich for a set of microorganisms with a particular phenotypic property, resulting in a partially purified, mixed culture)

If so, provide the taxonomic lineage as far as you have determined it and include the clone identifier. We will accept either of the following for organisms obtained from an enrichment culture: (See Box 11)

B. Were the sequences you want to submit extracted from:

1. Bulk environmental DNA (using universal primers)?

DNA that is PCR-amplified directly from source/host DNA (e.g. soil, ocean water, etc.) using universal primers

If so, provide the taxonomic lineage as far as you have determined it and include the clone identifier in the appropriate source modifier. Do not include the clone within the organism name itself. You must also include the isolation source modifier, within which you will give the specific environmental conditions from which the sample was isolated (soil, ocean water, etc.). We will accept any of the following for the culture lineage: (See Box 12)

2. Bulk environmental DNA (using species-specific primers, not gene-specific primers)?

DNA that is PCR-amplified directly from source/host DNA (e.g. soil, ocean water, etc.) using species-specific primers.

If so, provide the full binomial (genus species) name and include the isolate identifier in the appropriate source modifier. Do not include the isolate within the organism itself. You must also include the isolation source modifier giving the specific environmental conditions from which the sample was isolated (soil, ocean water, etc). Please include a note indicating that the sequence was amplified with species-specific primers. (See Box 13).

Is it OK for me to use BLAST results to identify the source organism from which I isolated my sequences?

Use BLAST scores only as a rough guide to the identity of the source organisms from which sequences were derived.

Assign source organisms to the same taxonomic rank as the best BLAST hits only at the genus level or higher, depending on the consistency among the best hits:

- Assign the source organism to a single genus if the best scores involve taxa from that genus
- Name the source organism using the next highest rank (family, order, class, or phylum) that includes all the best hits if there is inconsistency among the best hits as to the genus identity
- The source organism should be assigned the name 'bacterium <strain>' (for domain Bacteria) OR 'archaeon <strain>' (for domain Archaea) if there is uncertainty as to the proper phylum

Examine the "taxonomy reports" that are provided with the BLAST results. The initial report presents the information in the definition field of the associated sequence records.

While this field should be updated if there is a change in the taxonomic name or lineage of the source organism in that record, the individual, complete sequence record should be examined to be certain that the proper taxonomic information is utilized in the definition field.

Note: species names can be assigned to source organisms only if species-specific primers were used during amplification. Otherwise, you must use genus level or higher ranks to name the organism.

Box 10.

Format	Example
Genus species	Escherichia coli
Genus sp. strain identifier	Escherichia sp. 1234
Family bacterium strain identifier	Enterobacteriaceae bacterium 1234

Box 11.

Format	Example
Genus sp. enrichment culture clone 1234 clone identifier	Escherichia sp. enrichment culture clone
Family bacterium enrichment culture clone culture clone 1234 clone identifier	Enterobacteriaceae bacterium enrichment

Box 12.

Format	Example
uncultured Genus sp.	uncultured Escherichia sp.
uncultured Family bacterium	uncultured Enterobacteriaceae bacterium

Box 12 continues on next page...

Box 12 continued from previous page.

uncultured Kingdom

uncultured bacterium

Box 13.

Format

Full binomial (genus species) name

Example

Escherichia coli