

The design of source modifiers in GenBase is inspired by [source modifiers table](#) from GenBank database in NCBI. We inherit modifiers from GenBank and add our specific modifiers. The Microsoft xlsx template is provided to the users in the submission page. It is also provided with tips and control vocabulary.

User should fill in the “Source Modifiers Table” sheet in the template file. If the value of modifier is controlled, you can find the controlled vocabulary of the value in the worksheet named control word.

The Modifiers can be divided into three parts,

- Required Modifiers
 - GenBase has some mandatory items, such as Sequence_ID and Collection_date. This design allows users to fill in the source modifiers table as flexibly as possible. Please note that the Sequence_ID must correspond to the SeqID in FASTA file and cannot be less than or greater than.
- Optional but recommended
 - These modifiers are recommended by GenBase, which you can uniquely identify your samples from the same organism with them.
- Optional
 - These modifiers are optional.

Description of each modifier are listed below:

Type	Modifiers	Description
Required	Sequence_ID	Seqid for each sequence
Required	Collection_date	Date the specimen was collected.
Optional but recommended	Clone	Name of clone from which sequence was obtained, typically an alphanumeric ID.
Optional but recommended	Country/Region	The country where the sequence's organism was located. May also be an ocean or major sea. Additional region or locality information must be after the country name and separated by a ':'. For example: USA: Riverview Park, Ripkentown, MD. Validated values are available at the control words sheet.
Optional but recommended	Host	When the sequence submission is from an organism that exists in a symbiotic, parasitic, or other special relationship with some second organism, the 'host' modifier can be used to identify the name of the host species.
Optional but recommended	Host_sex	Sex of Host
Optional but recommended	Host_age	Age of Host

Optional but recommended	Isolate	Individual isolate from which the sequence was obtained, typically an alphanumeric sample ID.
Optional but recommended	Isolation_source	Describes the physical, environmental and/or local geographical source of the biological sample from which the sequence was derived.
Optional but recommended	Specimen_voucher	An identifier of the individual or collection of the source organism and the place where it is currently stored, usually an institution. This should be provided using the following format 'institution-code:collection-code:specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection-code is provided. Examples: 99-SRNP or UAM:Mamm:52179 or personal collection:Joe Smith:99-SRNP
Optional but recommended	Strain	Strain of organism from which sequence was obtained. For microbial records, the strain is an alphanumeric identifier that may be designated in any manner, for example, it may be based on the name of an individual or locality. As an example, for Escherichia coli K12, "K12" is the strain name/identifier.
Optional	Organelle/Location	The default Location for all sequences is 'Genomic'. If the sequence is not genomic, select the alternative organelle/location from the control words sheet. Details can be found here .
Optional	Altitude	Altitude in meters above or below sea level of where the sample was collected.
Optional	Authority	The author or authors of the organism name from which sequence was obtained.
Optional	Bio_material	An identifier for the biological material from which the nucleotide sequence was obtained, with optional institution code and collection code for the place where it is currently stored. This should be provided using the following format 'institution-code:collection-code:material_id'. material_id is mandatory, institution-code and collection-code are optional; institution-code is mandatory when collection-code is present. This qualifier should be used to annotate the identifiers of material in biological collections which include zoos and aquaria, stock centers, seed banks, germplasm repositories and DNA banks.

Optional	Biotype	Variety of a species (usually a fungus, bacteria, or virus) characterized by some specific biological property (often geographical, ecological, or physiological). Same as biotype.
Optional	Biovar	See biotype
Optional	Breed	The named breed from which sequence was obtained (usually applied to domesticated mammals).
Optional	Cell_line	Cell line from which sequence was obtained.
Optional	Cell_type	Type of cell from which sequence was obtained.
Optional	Chemovar	Variety of a species (usually a fungus, bacteria, or virus) characterized by its biochemical properties.
Optional	Collected_by	Name of person who collected the sample.
Optional	Cultivar	Cultivated variety of plant from which sequence was obtained.
Optional	Culture_collection	Institution code and identifier for the culture from which the nucleotide sequence was obtained, with optional collection code. This should be provided using the following format 'institution-code:collection-code:culture-id'. culture-id and institution-code are mandatory. This qualifier should be used to annotate live microbial and viral cultures, and cell lines that have been deposited in curated culture collections.
Optional	Dev_stage	Developmental stage of organism.
Optional	Ecotype	The named ecotype (population adapted to a local habitat) from which sequence was obtained (customarily applied to populations of <i>Arabidopsis thaliana</i>).
Optional	Forma	The forma (lowest taxonomic unit governed by the nomenclatural codes) of organism from which sequence was obtained. This term is usually applied to plants and fungi.
Optional	Forma_specialis	The physiologically distinct form from which sequence was obtained (usually restricted to certain parasitic fungi).
Optional	Fwd_primer_name	Name of forward PCR primer
Optional	Fwd_primer_seq	Nucleotide sequence of forward PCR primer
Optional	Genotype	Genotype of the organism.
Optional	Haplogroup	Name for a group of similar haplotypes that share some sequence variation
Optional	Haplotype	Haplotype of the organism.

Optional	Identified_by	Name of the person or persons who identified by taxonomic name the organism from which the sequence was obtained
Optional	Lab_host	Laboratory host used to propagate the organism from which the sequence was obtained.
Optional	Lat_Lon	Latitude and longitude, in decimal degrees, of where the sample was collected.
Optional	Note	Any additional information that you wish to provide about the sequence.
Optional	Pathovar	Variety of a species (usually a fungus, bacteria or virus) characterized by the biological target of the pathogen. Examples include <i>Pseudomonas syringae</i> pathovar tomato and <i>Pseudomonas syringae</i> pathovar tabaci.
Optional	Pop_variant	Name of the population variant from which the sequence was obtained
Optional	Rev_primer_name	Name of reverse PCR primer
Optional	Rev_primer_seq	Nucleotide sequence of reverse PCR primer
Optional	Segment	Name of viral or phage segment sequenced
Optional	Serogroup	Variety of a species (usually a fungus, bacteria, or virus) characterized by its antigenic properties. Same as serogroup and serovar.
Optional	Serotype	See Serogroup
Optional	Serovar	See Serogroup
Optional	Sex	Sex of the organism from which the sequence was obtained.
Optional	Sub_species	Subspecies of organism from which sequence was obtained.
Optional	Subclone	Name of subclone from which sequence was obtained.
Optional	Subtype	Subtype of organism from which sequence was obtained.
Optional	Substrain	Sub-strain of organism from which sequence was obtained.
Optional	Tissue_lib	Tissue library from which the sequence was obtained.
Optional	Tissue_type	Type of tissue from which sequence was obtained.
Optional	Type	Type of organism from which sequence was obtained.
Optional	Variety	Variety of organism from which sequence was obtained.
Optional	Comments	Additional comments for each Sequence

