

# Field definitions

Attribute names can vary greatly, especially those used to describe samples. We map all similar fields into a single representation, and normalise the format of the data wherever possible. To assist determining which fields are of interest in building search queries and result reports, a full listing of all fields and their definitions are detailed below.

Field	Description
accession	Accession number
allele	Name of the allele for the given gene
altitude	Altitude (in metres)
analysis_accession	Analysis accession number
analysis_alias	Submitter's name for the analysis
analysis_title	Brief analysis description
analysis_type	Type of sequence analysis
anticodon	Location of the anticodon of tRNA and the amino acid for which it codes
artificial_location	Indicates location is modified to adjust for the presence of a frameshift or internal stop codon
assembly_level	Assembly level (contig, scaffold, chromosome, complete genome)
assembly_name	Genome assembly name. When searching, considers all all live versions of the assembly. When returning, the name of the latest version.
assembly_title	Brief assembly description
base_count	Number of base pairs
bio_material	Identifier for biological material including institute and collection code
breed	Breed
broker_name	Name of broker for the submission
cell_line	Cell line from which the sample was obtained
cell_type	Cell type from which the sample was obtained
center_name	Submitting centre
checklist	Accession of ENA sample checklist
codon_start	Indicates the offset of the first complete codon relative to the first base of the coding feature
collected_by	Name of the person who collected the specimen

Field	Description
collecting_institute	Name of the institution to which the person collecting the specimen belongs. Format: Institute Name, Institute Address
collection_date	Date that the specimen was collected
country	Locality of sample isolation: country names, oceans or seas, followed by regions and localities
cram_index_aspera	Aspera links for CRAM index files
cram_index_ftp	FTP links for CRAM index files
cram_index_galaxy	Galaxy links for CRAM index files
cultivar	Cultivar (cultivated variety) of plant from which sample was obtained
culture_collection	Identifier for the sample culture including institute and collection code
dataclass	Sequence data class
depth	The distance below the surface of the water at which a measurement was made or a sample was collected (in metres)
description	Brief sequence description
dev_stage	Sample obtained from an organism in a specific developmental stage
ec_number	Enzyme commission number for enzyme product of sequence
ecotype	A population within a given species displaying traits that reflect adaptation to a local habitat
elevation	The elevation of the sampling site as measured by the vertical distance from mean sea level (in metres)
embl_file	Flat file for the set
environment_biome	Environment (biome). Biomes are defined based on factors such as plant structures, leaf types, plant spacing, and other factors like climate. Examples include: desert, taiga, deciduous woodland, or coral reef
environment_feature	Environmental feature level includes geographic environmental features. Examples include: harbor, cliff, or lake
environment_material	The environmental material level refers to the material that was displaced by the sample, or material in which a sample was embedded, prior to the sampling event. Examples include: air, soil, or water
environmental_package	MIGS/MIMS/MIMARKS extension for reporting of measurements and observations obtained from one or more of the environments where the sample was obtained
environmental_sample	Identifies sequences derived by direct molecular isolation from an environmental DNA sample
event_label	Label given to sampling event
exception	Indicates that the coding region cannot be translated using standard biological rules

Field	Description
experiment	A brief description of the nature of the experimental evidence
experiment_accession	Experiment accession number
experiment_alias	Submitter's name for the experiment
experiment_title	Brief experiment title
experimental_factor	Experimental factors are essentially the variable aspects of an experiment design which can be used to describe an experiment, or set of experiments, in an increasingly detailed manner
fasta_file	FASTA file for the set
fastq_aspera	Aspera links for FASTQ files
fastq_bytes	Size (in bytes) of FASTQ files
fastq_ftp	FTP links for FASTQ files
fastq_galaxy	Galaxy links for FASTQ files
fastq_md5	MD5 checksum of FASTQ files
first_public	Date when made public
function	Function attributed to a sequence
gene	Symbol of the gene corresponding to a sequence region
gene_synonym	Synonymous, replaced, obsolete or former gene symbol
genome_representation	Whether the genome assembly is a full or partial genome
geo_accession	GEO accession
germline	Indicates whether the sample is an unrearranged molecule that was inherited from the parental germline
haplotype	Combination of alleles that are linked together on the same physical chromosome
host	Natural (as opposed to laboratory) host to the organism from which sample was obtained
host_body_site	Name of body site where the sample was obtained from, such as a specific organ or tissue
host_common_name	Common name of the natural host organism from which the sample was obtained
host_genotype	Genotype of the host
host_gravidity	Whether or not subject is gravid, including date due or date post-conception where applicable
host_growth_conditions	Literature reference giving growth conditions of the host
host_phenotype	Phenotype of the host
host_scientific_name	Scientific name of the natural (as opposed to laboratory) host to the organism from which sample was obtained

Field	Description
host_sex	Physical sex of the host
host_status	Condition of host (eg. diseased or healthy)
host_subject_id	A unique identifier by which each subject can be referred to, de-identified
host_tax_id	NCBI taxon ID of the host
identified_by	Name of the taxonomist who identified the specimen
inference	A structured description of non-experimental evidence
influenza_test_method	Method by which the current assessment of a sample as flu positive/negative is made
influenza_test_result	Classification of a sample as flu positive or negative based on the test performed and reported
instrument_model	Instrument model used in sequencing experiment
instrument_platform	Instrument platform used in sequencing experiment
investigation_type	The study type targeted by the sequencing
isolate	Individual isolate from which sample was obtained
isolation_source	Describes the physical, environmental and/or local geographical source of the sample
keywords	Keywords associated with sequence
lab_host	Scientific name of the laboratory host used to propagate the source organism for the sample
last_updated	Date when record was last updated
library_layout	Sequencing library layout
library_name	Sequencing library name
library_selection	Method used to select or enrich the material being sequenced
library_source	Source material being sequenced
library_strategy	Sequencing technique intended for the library
location	Geographic location of isolation of the sample. Latitude and longitude are given in decimal degrees. When using latitude and longitude in the geospatial search functions, positive and negative values should be given to represent direction. When returned in the results report, N/S and E/W are displayed with latitude and longitude.
locus_tag	A submitter-supplied, systematic, stable identifier for a gene and its associated features
map	Map position of feature
marine_region	Geographical origin of the sample as defined by the marine region
marker	Marker classification

Field	Description
master_file	Flat file for the set master
mating_type	Mating type of the organism from which the sequence was obtained
mol_type	in vivo molecule type of the sequence
nominal_length	Average fragmentation size of paired reads
old_locus_tag	Deprecated submitter-supplied, systematic, stable identifier for a gene and its associated features
operon	Name of the group of contiguous genes transcribed into a single transcript
organelle	Membrane-bound intracellular structure from which the sequence was obtained
other_pathogens_result	Classification of a sample as positive or negative based on the test performed and reported
other_pathogens_tested	Classification of pathogenic organisms other than influenza virus tested in the current assessment of a sample
parent_accession	Parent sequence accession number
parent_study	Parent study accession number
ph	pH measurement
plasmid	Name of naturally occurring plasmid from which the sequence was obtained
product	Name of the product associated with the feature
project_name	Name of the project within which the sequencing was organised
protein_id	A stable protein identifier issued by INSDC
protocol_label	The protocol used to produce the sample
pseudo	Indicates whether the feature is non-functional
pseudo_gene	Indicates that this feature is a pseudogene
read_count	Number of reads
receipt_date	Date on which the sample was received
region	Geographical origin of the sample as defined by the specific region name followed by the locality name
ribosomal_slippage	Indicates ribosomal slippage (change to an alternative reading frame) during protein translation
rna_class	Classification of RNA
run_accession	Run accession number
run_alias	Submitter's name for the run
salinity	Salinity of water at the time of taking the sample

Field	Description
sample_accession	Sample accession number
sample_alias	Submitter's name for the sample
sample_collection	The method or device employed for collecting the sample
sample_title	Brief sample title
sampling_campaign	The activity within which this sample was collected
sampling_platform	The large infrastructure from which this sample was collected
sampling_site	The site/station where this sample was collection
scientific_name	Scientific name of the organism from which the sample was derived
secondary_sample_accession	Secondary sample accession number
secondary_study_accession	Secondary study accession number
sequence_md5	MD5 checksum of the sequence
sequencing_method	Sequencing method used
serovar	Serological variety of a species characterized by its antigenic properties
set_files	Flat file for the set
sewage_type	Type of sewage based on origin
sex	Sex of the organism from which the sample was obtained
specimen_voucher	Identifier for the sample culture including institute and collection code
sra_aspera	Aspera links for NCBI SRA format files
sra_bytes	Size (in bytes) of NCBI SRA format files
sra_ftp	FTP links for NCBI SRA format files
sra_galaxy	Galaxy links for NCBI SRA format files
sra_md5	MD5 checksum of NCBI SRA format files
standard_name	Accepted standard name for a feature
strain	Strain from which sample was obtained
study_accession	Study accession name
study_alias	Submitter's name for the study
study_description	Detailed sequencing study description
study_title	Brief sequencing study description
sub_species	Name of sub-species of organism from which sample was obtained
sub_strain	Name or identifier of a genetically or otherwise modified strain from which sample was obtained

Field	Description
submission_accession	Submission accession number
submitted_aspera	Aspera links for submitted files
submitted_bytes	Size (in bytes) of submitted files
submitted_format	Format of submitted reads
submitted_ftp	FTP links for submitted files
submitted_galaxy	Galaxy links for submitted files
submitted_host_sex	Physical sex of the host as provided by the submitter. This is in contrast to the “host_sex” field which contains standardised values.
submitted_md5	MD5 checksum of submitted files
submitted_sex	Sex of the organism from which the sample was obtained. This is in contrast to the “sex” field which contains standardised values.
target_gene	Targeted gene or locus name for marker gene studies
tax_division	Taxonomic division of the organism from which the sample was obtained
tax_id	NCBI taxon ID of the organism from which the sample was obtained
taxonomy	A virtual field representing NCBI taxonomic classification. Searchable using one of the taxonomy functions
temperature	Temperature of the sample at time of sampling
tissue_lib	Tissue library from which sample was obtained
tissue_type	Tissue type from which the sample was obtained
topology	Sequence topology: circular or linear
trans_splicing	Indicates exons from two RNA molecules are ligated in intermolecular reaction to form mature RNA
transl_except	A single codon translation that does not conform to genetic code
transl_table	Indicates the genetic code table used if other than universal genetic code table
variety	Variety (varietas, a formal Linnaean rank) of organism from which sample was derived