

1. General sample information

1.1. *Sample*: sample's general data

campo	tipo	descripción
name	TEXTOLIBRE	Sample name
stype	TEXTOLIBRE	Sample type. Description of the sample
ssize	REAL	Sample size (weight, volume)
ssizeunit	VALOR A ESCOGER	Units of <i>ssize</i> Complementary value for the sample size value (<i>ssize</i>) <ul style="list-style-type: none">• μl• ml• l• mg• g• kg• ...
keywords	VALORES A ESCOGER	Keywords The user will be able to choose the value from a list, which can be expanded with new additions.

1.2. *Metadata on the origin of the sample*: Sampling conditions and contextual information related to sample collection.

campo	tipo	descripción
coun	TEXTOLIBRE	Country of origin
regi	TEXTOLIBRE	Region of origin
geol	TEXTOLIBRE	Geographic location
lati	REAL	Latitude
long	REAL	Longitude
alti	REAL	Altitude in meters, over level at sea, where the sample was taken.
dept	REAL	Depth, in meters, where the sample was taken, with respect to the altitude. That is, if the sample was taken in a lake placed on a mountain, the altitude of the lake's surface is the reference point to measure the depth into the lake where the sample was taken.
sour	TEXTOLIBRE	Other relevant information about the sample
dat	DATE	Sampling date
hocs	TIME	Hour of collection of the sample [sharp hours]. Exact time at which the sample was collected. This is the local time, serving as a reference for potential later considerations.
dati	DATE	Date of isolation
tems	REAL	Temperature of the sample [in celsius degrees]
phsa	REAL	pH of the sample
sals	REAL	Salinity of the sample [in percentage %]

2. NON-CLEARLY CLASSIFIED ITEMGROUPS

2.1. Other information details on the organism:

ATENCIÓN ¿Esto es en general? ¿Se aplica solo si se ha identificado un organismo?

Campo	tipo	descripción
bios	VALOR A ESCOGER	Biosafety level Possible values: <ul style="list-style-type: none">• BSL-1• BSL-2• BSL-3• BSL-4
bior	TEXTO LIBRE	Stable known association with other organisms
host	TEXTO LIBRE	Organism with which does it interact
path	TEXTO LIBRE	Known pathogenicity.
extr	TEXTO LIBRE	Miscellaneous, extraordinary features relevant for the description

2.2. Nagoya protocol metadata

Atención: falta identificar qué apartados pueden estar relacionados con el protocolo de Nagoya.

campo	tipo	descripción
nagoya	TEXTO LIBRE	Information related to the Nagoya Protocol ¡Atención! ¿Un comentario, o algo más específico? Normalmente es un código ¿hay una lista de códigos?

3. EXPERIMENTS.

Upon the sample, several actions can be done. Let's name them Experiment (following ENA nomenclature) another possible name for this section is TREATMENTS.

MORE THAN ONE EXPERIMENT/TREATMENT can be applied to a sample?

3.1. Information on the kind of material used for the description:

Description about how the sample has been obtained and processed.

campo	tipo	descripción
cult	VALOR A ESCOGER	Is the sample cultured or uncultured? <ul style="list-style-type: none"> • cultured • uncultured
koma	VALOR A ESCOGER	Kind of material. Depends on the value of cult: Possible values <ul style="list-style-type: none"> • If cultured <ul style="list-style-type: none"> ◦ Isolated bacteria ◦ Isolated archaea ◦ Isolated virus ◦ Isolated Eukaryote ◦ Enrichment • If uncultured <ul style="list-style-type: none"> ◦ MAG ◦ SAG ◦ Virus contig/genoma ◦ Virome ◦ Transcriptome ◦ Exome ◦ Epigenome ◦ Metabolome ◦ Meta-epigenome ◦ Metagenome ◦ Metaproteome ◦ Metatranscriptome ◦ Targeted Sequencing ◦ Plasmids and Mobile Genetic Elements ◦ Gene ◦ other→
coth	TEXTO LIBRE	Other comments about the sample

The following values seem that are related to the identification of a new taxon. So, are they related to section 2.1?

typn	SÍ/NO	Is this the designed type material for a new taxon?
seqcode	TEXTO LIBRE	Sequencing code
txnr	ENTERO	NCBI Taxonomy number of the type material (in case of member of an extant taxon)
ccsu	SÍ/NO	If the sample is cultured, has been submitted to culture collection?

type	TEXTO LIBRE	If TYPN is true, designated type strain name/MAG/SAG
coln	TEXTO LIBRE	Strain collection numbers
strccol	ENTERO	Strain collection number Identificador (NCBI?) De la colección donde se encuentra la cepa

3.2. Basic metabolic traits determined under culture conditions (or inferred when possible):

datos metabólicos que se pueden deducir de alguna manera a partir de un cultivo.

used only if the field `cult` (in section 3.1) has been defined as `cultured`

campo	tipo	descripción
emet	TEXTO LIBRE	Metabolic category Possibly an enumerated value (VALOR A ESCOGER) What are the values?
orel	VALOR A ESCOGER	Relationship with O ₂ Possible values: <ul style="list-style-type: none"> • Aerobic metabolism • Anaerobic metabolism <ul style="list-style-type: none"> ◦ Obligate ◦ Facultative • Microaerophiles • Aerotolerant anaerobes
elac	TEXTO LIBRE	Terminal electron acceptor. Chemical compound that serves to receive electrons in cellular respiration, photosynthesis, fermentation processes...
temo	REAL	Temperature optimum [in degrees Celsius]
teml	REAL	Lowest temperature for growth [in degrees Celsius]
temh	REAL	Highest temperature for growth [in degrees Celsius]
temc	VALOR A ESCOGER	Temperature category (value related to temo) Possible Values: <ul style="list-style-type: none"> • Psychrophyles (-20°C, 10°C] • Mesophiles (10°C, 45°C] • Moderate thermophiles (45°C, 60°C] • Thermophiles (60°C, 80°C] • Hyperthermophiles (80°C, 122°C)
phop	REAL	pH optimum
phlo	REAL	Lowest pH for growth
phhi	REAL	Highest pH for growth

campo	tipo	descripción
phca	VALOR A ESCOGER	<p>pH category (value related to phop)</p> <p>Possible values:</p> <ul style="list-style-type: none"> • Ultra acidic ($-\infty$, 3.5] • Extremely acidic (3.5, 4.5] • Very strongly acidic (4.5, 5.0] • Strongly acidic (5.0, 5.5] • Moderately acidic (5.5, 6.0] • Slightly acidic (6.0, 6.5] • Neutral (6.5, 7.3] • Slightly alkaline (7.3, 7.8] • Moderately alkaline (7.8, 8.4] • Strongly alkaline (8.4, 9.0] • Very strongly alkaline (9.0, 10.5] • Hyper alkaline (10.5, $+\infty$)
salo	REAL	Salinity optimum (0%-100%) [in percentage]
sall	REAL	Lowest NaCl concentration for growth [in percentage]
salh	REAL	Highest NaCl concentration for growth [in percentage]
salw	TEXTO LIBRE	Other salts besides NaCl to be reported
salc	VALOR A ESCOGER	<p>Salinity category(value related to salo)</p> <p>Possible values:</p> <ul style="list-style-type: none"> • Halotolerants [0%, 1%) • Slight halophiles [1%, 3%) • Moderate halophiles [3%, 15%) • Extreme halophiles [15%, 32%]

3.2. Gene and genome information

used only if the field `cult` (in section 3.1) has been defined as `uncultured`

campo	tipo	descripción
gsta	Valor a escoger	Genome current status. Possible values: <ul style="list-style-type: none"> • complete • partial • draft
seqdepth	FLOAT	Sequencing depth (Isolates/MAG) Número de veces que un nucleótido es secuenciado: Número de bases secuenciadas/tamaño del genoma Falta aclariment
sixteensr	TEXTO LIBRE	16S rRNA gene accession number
twentythreesr	TEXTO LIBRE	23S rRNA gene accession number
hkgn	VALORES A ESCOGER	Alternative housekeeping genes. List of values from a table. The user can add more values as needed.
meca	TEXTO LIBRE	INSDC (International Nucleotide Sequence Database Collaboration) metagenome accession number.
gare	TEXTO LIBRE	genome/MAG/SAG accession number [RefSeq/ENA]
bin	TEXTO LIBRE	genome/MAG/SAG accession number [other] Código de acceso en formato diferente a RefSeq o EMBL
completeness	REAL	Completeness (0%-100%) [in percent]. Percentage of the genome already processed.
contamination	REAL	Level of contamination (0%-100%) [in percent]. Proportion of reads not belonging to MAGs detected.
method	VALOR A ESCOGER	Method used to estimate the integrity and contamination of the sample. Possible values: <ul style="list-style-type: none"> • BUSCO • CheckM • CheckV • OrthoDB • Kraken • BlobTools • Anvi'o • MiGA • Other → to add new values The user can select one of the values or add a new one, which will be added to the list

campo	tipo	descripción
gsiz	ENTERO	Estimated (for incomplete) or actual (for complete) genome/metagenome size. Unit bp (base pairs)
ggcm	REAL	Molar percentage of Guanine and Cytosine in DNA.
asem	VALOR A ESCOGER	<p>Software used for assembly</p> <p>Possible values:</p> <ul style="list-style-type: none"> • SPAdes • Velvet • ABySS • Canu • Flye • Shasta • MaSuRCA • HybridSPAdes • MEGAHIT • IDBA-UD • MetaSPAdes • Trinity • Oases • SOAPdenovo • NOVOPlasty • Unicycler • OTHERS → to add new values <p>The user can select one of the values or add a new one, which will be added to the list</p>
asftparams	TEXTO LIBRE	Parameters used with asem

campo	tipo	descripción
seqt	VALOR A ESCOGER	<p>Sequencing technology used</p> <p>Possible values:</p> <ul style="list-style-type: none"> • FastQC • Trimmomatic • Cutadapt • SPAdes • MEGAHIT • IDBA-UD • BWA • Bowtie2 • HISAT2 • GATK • FreeBayes • Samtools • Prokka • InterProScan • EggNOG-mapper • MetaPhlAn • Kraken2 • MetaGeneMark • OTHERS → to add new values <p>The user can select one of the values or add a new one, which will be added to the list</p>
bins	VALOR A ESCOGER	<p>Binning software used</p> <p>Possible values:</p> <ul style="list-style-type: none"> • MetaBAT • MaxBin • CONCOCT • DAS Tool • OTHERS → to add new values <p>The user can select one of the values or add a new one, which will be added to the list</p>
binsparams	TEXTO LIBRE	Params used with bins

3.3. Sequencing

Additional data related to sequencing.

Atención ¿No será mejor determinar primero lo relacionado con raw reads y establecer experiments/treatments separados en función de si se han determinado trimmed reads se ha realizado un ensamblado o si se han predicho genes?

campo	tipo	descripción
target	VALOR A ESCOGER	Target nucleic acids Possible values: <ul style="list-style-type: none">• DNA• RNA
rreadacc	ENTERO	Raw reads accession code
rreadaccref	VALOR A ESCOGER	Raw reads reference Database. Database with respecto to rreadacc has its value. <ul style="list-style-type: none">• NCBI•
rreads	TEXTO LIBRE	File containing the raw reads
rreadsnum	ENTERO	Number of raw reads.
rreadsbp	ENTERO	Number of base pairs of raw reads
treads	TEXTO LIBRE	File containing the trimmed reads
treadsnum	ENTERO	Number of trimmed reads
treadsbp	ENTERO	Number of base pairs of trimmed reads.
coverage	FLOAT	Coverage (Nonpareil) Proporción de "target region" que se ha secuenciado al menos una vez. <i>Nonpareil</i> es un software específico
assembled	TEXTO LIBRE	File containing the assembly
asize	ENTERO	Assembly size in base pairs
contignumber	ENTERO	Number of contigs
pgenes	TEXTO LIBRE	File containing the predicted genes