ESTÁNDAR PARA DATOS DE SECUENCIAS DE ADN (GenBank)

FECHA DE ACTUALIZACIÓN: FEBRERO 2020

DESCRIPCIÓN DEL ESTÁNDAR

GenBank es la base de datos de secuencias genéticas de la National Institutes of Health (NIH), una colección de todas las secuencias de ADN disponibles (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank forma parte de la International Nucleotide Sequence Database Collaboration (INSDC), que comprende el DNA DataBank de Japón (DDBJ), el European Nucleotide Archive (ENA), y GenBank del National Center for Biotechnology Information (NCBI). Estas tres organizaciones intercambian datos diariamente (NIH, 2017).

Las secuencias de ácidos nucleicos proporcionan el punto de partida fundamental para describir y comprender la estructura, función y desarrollo de organismos genéticamente diversos. Los bancos de datos de secuencias de ácidos nucleicos de GenBank, EMBL (European Molecular Biology Laboratory) y DDBJ han utilizado desde su inicio, tablas de sitios y características para describir los roles y ubicaciones de los dominios de secuencia de orden superior y los elementos dentro del genoma de un organismo (INSDC, 2017).

En febrero de 1986, GenBank y EMBL comenzaron un esfuerzo de colaboración (la DDBJ se unió en 1987) para diseñar un formato de tabla de características comunes y estándares comunes para la práctica (INSDC, 2017).

JUSTIFICACIÓN DE SU USO

El objetivo general del diseño del diccionario de datos (características) es proporcionar un amplio vocabulario para describir las características de las secuencias y su contexto para brindar al usuario una estructura que facilite su manejo. La documentación de la “The DDBJ/ENA/GenBank Feature Table Definition” representa las reglas compartidas que permiten que las tres fuentes de datos (DDBJ, ENA y GenBank) intercambien información de manera cotidiana (INSDC, 2017).

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

DICCIONARIO

El diccionario se divide en tres secciones:

1. “Lista de campos en formato GenBank”: contiene los campos principales que se presentan en una ficha de dicha base de datos
2. “Lista de “Feature keys” y su descripción”: contiene las características del campo FEATURES de la “Lista de campos en formato GenBank”
3. “Lista de “Qualifiers” y su descripción”: contiene los calificadores que brindan una descripción más específica a los “Feature keys”

LISTA DE CAMPOS EN FORMATO GENBANK

Término	Definición
1. LOCUS	Contiene varios elementos como: el nombre del locus, longitud de la secuencia, tipo de molécula, la división según GenBank y la fecha de modificación (p. ej., JF412835 668 bp DNA linear ENV 01-JUL-2012).
1.1. Locus Name	La regla que debe cumplir este campo es que el nombre del locus sea único. Tal es el caso de aquellos registros que tienen 6 caracteres (p. ej. U12345), el nombre del locus es la primera letra del género y epíteto específico, seguido del número de acceso. Para aquellos con 8 caracteres (p. ej. AF123456), el nombre del locus es el número de acceso.
1.2. Sequence Length	Número de pares de bases de nucleótidos (o residuos de aminoácidos) en la secuencia registrada. No hay un máximo en el tamaño de la secuencia que se pueda someter a GenBank, pero sí un mínimo de 50 pares de bases.
1.3. Molecule Type	El tipo de molecula que fue secuenciada. Cada registro de GenBank debe contener datos de secuencia continua de un mismo tipo de molécula. Se recomienda el uso de vocabulario controlado: “genomic DNA”, “genomic RNA”, “precursor RNA”, “mRNA (cDNA)”, “ribosomal RNA”, “transfer RNA”, “small nuclear RNA”, y “small cytoplasmic RNA”.

Término	Definición
1.4. GenBank Division	La división según GenBank a la que pertenece el registro, se indica con tres letras. Se recomienda el uso de vocabulario controlado: 1. PRI - primate sequences 2. ROD - rodent sequences 3. MAM - other mammalian sequences 4. VRT - other vertebrate sequences 5. INV - invertebrate sequences 6. PLN - plant, fungal, and algal sequences 7. BCT - bacterial sequences 8. VRL - viral sequences 9. PHG - bacteriophage sequences 10. SYN - synthetic sequences 11. UNA - unannotated sequences 12. EST - EST sequences (expressed sequence tags 13. PAT - patent sequences 14. STS - STS sequences (sequence tagged sites) 15. GSS - GSS sequences (genome survey sequences) 16. HTG - HTG sequences (high-throughput genomic sequences) 17. HTC - unfinished high-throughput cDNA sequencing 18. ENV - environmental sampling sequences Algunas de las divisiones tienen secuencias de grupos específicos de organismos, mientras que otras (EST, GSS, HTG, etc.) contienen datos generados por tecnologías específicas de secuenciación de organismos diferentes. Las divisiones son históricas y no reflejan la taxonomía actual del NCBI. En su lugar, sirven como una forma de dividir GenBank en pequeñas partes para aquellos que desean enviar por FTP la base de datos.
1.5. Modification Date	La fecha de última modificación del campo LOCUS. En algunos casos, puede corresponder a la fecha de publicación del registro, pero no hay manera de saberlo.
2. DEFINITION	Breve descripción de la secuencia; incluye información como el organismo, nombre del gen o proteína o alguna descripción de la función de la secuencia (si la secuencia es no codificante) (p. ej.,

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Término	**Definición**
	Uncultured Sphingomonadaceae bacterium clone 40_1_1 16S ribosomal RNA gene, partial sequence). Si la secuencia tiene una región codificante (CDS), la descripción puede ir seguida de un calificador como “complete cds”.
3. ACCESSION	Identificador único para el registro de la secuencia. Un número de acceso aplica para el registro completo y es usualmente una combinación de letras y números. Dicha combinación puede ser una letra seguida de cinco números (p. ej., U12345) o dos letras seguidas de seis números (p. ej., AF123456). Algunos números de acceso pueden ser más largos, dependiendo del tipo de secuencia que se registre. Los números de acceso no cambian aún si la información del registro es modificada por solicitud de los autores, únicamente se otorgará un nuevo número de acceso si los autores someten un nuevo registro que combine las secuencias previas o si reemplaza el registro previo.
4. VERSION	Número de identificación de la secuencia de nucleótidos que representa una secuencia única y específica dentro de la base de datos de GenBank. Este número de identificación utiliza el formato “accession.version”, implementado por GenBank / EMBL / DDBJ en febrero de 1999 (p. ej., AH001728.2, siendo el número de acceso AH001728 y la versión 2). Si se realiza algún cambio en los datos de la secuencia (incluso una única base), el número de versión se incrementará (p. ej., U12345.1 a U12345.2) pero el número de acceso se mantiene. El sistema de numerodeacceso.versión de los identificadores se ejecuta en paralelo al sistema de números GI, es decir, cuando se realiza cualquier cambio en una secuencia, recibe un nuevo número de GI y un incremento en el número de su versión.
4.1. GI	“GenInfo Identifier”, número de identificación de la secuencia, en este caso para la secuencia de nucleótidos. Si una secuencia registrada cambia, se asignará un nuevo número GI. También se asigna un número GI por separado a cada traducción de proteína dentro de un registro de secuencia de nucleótidos, y se asigna un nuevo GI si la traducción de proteína cambia.

Término	Definición
5. KEYWORDS	Palabra o frase que describe la secuencia. Si no se incluyen palabras clave, el campo debe contener solo un punto.
6. SOURCE	Información de formato libre que incluye una forma abreviada del nombre del organismo, a veces seguido de un tipo de molécula (p. ej., uncultured Sphingomonadaceae bacterium; Pinus taeda (loblolly pine); Cyprinella xanthicara (Cuatro Cienegas shiner); chloroplast Pinus ponderosa).
6.1. Organism	Nombre científico del organismo del que proviene la secuencia (género y epíteto específico, según corresponda) y su linaje, basado en el esquema de clasificación filogenética utilizado en la Base de atos Taxonómicos del NCBI. Si el linaje completo del organismo es muy largo, se mostrará abreviado en el registro de GenBank y el linaje completo estará disponible en la Base de datos Taxonómicos. Por ejemplo, uncultured phingomonadaceae bacterium Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; environmental samples.
7. REFERENCE	Publicaciones de los autores de la secuencia que discuten los datos reportados en el registro. Las referencias se ordenan automáticamente dentro del registro según la fecha de publicación, mostrando primero las referencias más antiguas.
7.1. Authors	Lista de autores en orden de aparición en el artículo citado. Por ejemplo, Pajares,S., Bonilla-Rosso,G., Travisano,M., Eguiarte,L.E. and Souza,V.
7.2. Title	Título del trabajo publicado o título provisional del no publicado. Algunas veces aparece “Direct Submission” en lugar del título del artículo. Esto es porque la última cita del campo “REFERENCE” tiende a contener información sobre el remitente de la secuencia, en lugar de una cita bibliográfica. La última cita se llama, “submitter block”. Información adicional se puede encontrar seguido del encabezado “Direct Submission”. Algunos registros anteriores no contienen el “submitter block”.
7.3. Journal	Abreviatura del nombre de la revista. Se puede obtener a través de Entrez de la base de datos: https://www.ncbi.nlm.nih.gov/nlmcatalog/journals

Término	Definición
7.4. Pubmed	Identificador de PubMed (PMID). Las referencias que incluyen ID de PubMed contienen enlaces desde el registro de la secuencia al registro de PubMed correspondiente. De la misma manera, los registros de PubMed que contienen números de acceso en el campo SI (secondary source identifier) contienen enlaces a los registros de la secuencia.
8. Direct Submission	Información de contacto del remitente, como instituto/departamento y dirección postal. Esta es siempre la última cita en el campo “REFERENCE”.
8.1. Authors	Nombre del remitente.
8.2. Title	Contiene “Direct Submission”.
8.3. Journal	Dirección postal del remitente.
9. FEATURES	Información sobre genes y productos génicos, así como regiones de importancia biológica reportadas en la secuencia. Estos pueden incluir regiones de la secuencia que codifican proteínas y moléculas de ARN, así como otras características. Existe una lista definida de características y su descripción (ver “Lista de “Feature keys” y su descripción”) que corresponde al “Appendix II: Feature keys reference” de la “The DDBJ/ENA/GenBank Feature Table Definition” (INSDC) (http://www.insdc.org/documents/feature_table.html#7.2).
9.1. Qualifiers	Los calificadores proporcionan un mecanismo para brindar información adicional sobre las características (“FEATURES”). El formato para presentar los calificadores es iniciar el renglón con una diagonal (/) seguida del nombre del calificador y, si aplica, un signo igual (=) y el valor. Cada calificador debe tener un solo valor, de ser necesario deberán presentarse en lista, p. ej.: /culture_collection="ATCC:11775" /culture_collection="CECT:515" Dado que los calificadores refieren distintos tipos de información, existen varios formatos en sus valores: • Texto libre: la mayoría de los valores de los calificadores serán una frase de texto descriptivo que debe escribirse entre comillas dobles. Si el texto ocupa más de un renglón, se requiere solo un par de comillas dobles. Si se

Término	Definición
	usan comillas dobles dentro del texto, este deberá escaparse colocando comillas dobles, p. ej.: /note="This is an example of ""escaped"" quotation marks" - Vocabulario controlado o valores numerados: algunos calificadores requieren valores de vocabulario controlado y se deben ingresar sin comillas. Este vocabulario controlado no distingue entre mayúsculas y minúsuclas, así que pueden mostrarse en cualquier combinación, p. ej.: /direction=Left; /direction=left; /direction=LEFT, todas estas variantes son aceptadas ya que transmiten el mismo significado. La lista del vocabulario (ver “Lista de “Qualifiers” y su descripción”) corresponde al “Appendix III: Summary of qualifiers for feature keys” de la “The DDBJ/ENA/GenBank Feature Table Definition” (INSDC) (http://www.insdc.org/documents/feature_table.html#7.3). - Cita o números de referencia: la cita o número de referencia publicado (como se enumera en la entrada “REFERENCE”), este valor deberá escribirse entre corchetes (p. ej., [3]) para distinguirlo de otros números. - Secuencias: secuencia literal de bases de nucleótidos, p. ej.: join(12..45,”atgcatt”, 988..1050).
10. ORIGIN	Este campo puede quedar vacío o asignarse la palabra “Unreported”, o puede contener un indicador del comienzo de la secuencia, generalmente involucrando un sitio de escisión de restricción determinado experimentalmente o el locus (si está disponible). Esta información está presente sólo en los registros más antiguos.

LISTA DE "FEATURE KEYS" Y SU DESCRIPCIÓN

La información obedece el siguiente formato y orden:

Feature Key: nombre de la característica.

Definition: descripción de la característica.

Mandatory qualifiers: calificadores obligatorios para la característica. Si no hay calificadores obligatorios, este campo se omite.

Optional qualifiers: calificadores opcionales asociados a la característica.

Organism scope: organismos válidos para la característica. Si aplica para cualquier organismo, entonces el campo se omite.

Molecule scope: tipos de moléculas válidas para la característica. Si aplica para cualquier tipo de molécula, entonces el campo se omite.

References: citas de los reportes publicados, usualmente respaldan la característica de la secuencia consenso.

Comment: comentarios y aclaraciones

Los "Feature keys" indican:

1. La naturaleza biológica de la característica
2. Información sobre cambios u otras versiones de la secuencia

Los "Feature keys" permiten al usuario recuperar rápidamente características similares o características con funciones equivalentes.

LISTA DE "QUALIFIERS" Y SU DESCRIPCIÓN

La información obedece el siguiente formato:

Qualifier: nombre del calificador, que requiere de un valor si contiene un signo igual.

Definition: descripción del calificador.

Value format: formato del valor, si es necesario.

Example: ejemplo del calificador con valor.

Comment: comentarios y aclaraciones.

LISTA DE "FEATURE KEYS" Y SU DESCRIPCIÓN

Feature key	**assembly_gap**
Definition	Gap between two components of a genome or transcriptome assembly.
Mandatory qualifiers	/ estimated_length=unknown or <integer> /gap_type="TYPE" /linkage_evidence="TYPE" (Note: Mandatory only if the /gap_type is "within scaffold" or "repeat within scaffold". If there are multiple types of linkage_evidence they will appear as multiple /linkage_evidence="TYPE" qualifiers. For all other types of assembly_gap features, use of the /linkage_evidence qualifier is invalid.). Mandatory qualifiers under assembly_gap feature for transcriptome shotgun assemblies (TSA): /estimated_length=<integer> /gap_type="within scaffold" and linkage_evidence="TYPE" where TYPE can not be "unspecified"
Comment	The location span of the assembly_gap feature for an unknown gap has to be specified by the submitter; the specified gap length has to be reasonable (less or = 1000) and will be indicated as "n"'s in sequence. However, the value for the estimated_length of assembly_gap features within a single (non-CON) transcriptome record must be an integer and can not be "unknown".
Feature key	**C_region**
Definition	Constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token)

	/map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Organism scope	eukaryotes
Feature Key	**CDS**
Definition	Coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature includes amino acid conceptual translation.
Optional qualifiers	/allele="text" /artificial_location="[artificial_location_value]" /citation=[number] /codon_start=<1 or 2 or 3> /db_xref="<database>:<identifier>" /EC_number="text" /exception="[exception_value]" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /number=unquoted text (single token) /old_locus_tag="text" (single token) /operon="text" /product="text" /protein_id="<identifier>" /pseudo /pseudogene="TYPE" /ribosomal_slippage

	/standard_name="text" /translation="text" /transl_except=(pos:<base_range>,aa:<amino_acid>) /transl_table =<integer> /trans_splicing
Comment	/codon_start has valid value of 1 or 2 or 3, indicating the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature; /transl_table defines the genetic code table used if other than the universal genetic code table; genetic code exceptions outside the range of the specified tables is reported in /transl_except qualifier; /protein_id consists of a stable ID portion (3+5 format with 3 position letters and 5 numbers) plus a version number after the decimal point; when the protein sequence encoded by the CDS changes, only the version number of the /protein_id value is incremented; the stable part of the /protein_id remains unchanged and as a result will permanently be associated with a given protein.
Feature Key	**centromere**
Definition	Region of biological interest identified as a centromere and which has been experimentally characterized.
Optional qualifiers	/citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /note="text" /standard_name="text"
Comment	The centromere feature describes the interval of DNA that corresponds to a region where chromatids are held and a kinetochore is formed.
Feature Key	**D-loop**
Definition	Displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA

	by a single stranded invader in the reaction catalyzed by RecA protein
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token)
Molecule scope	DNA
Feature Key	**D_segment**
Definition	Diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Organism scope	eukaryotes

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Feature Key	**exon**
Definition	Region of genome that codes for portion of spliced mRNA, rRNA and tRNA; may contain 5'UTR, all CDSs and 3' UTR
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /EC_number="text" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /number=unquoted text (single token) /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text" /trans_splicing
Feature Key	**gap**
Definition	Gap in the sequence.
Mandatory qualifiers	/estimated_length=unknown or <integer>
Optional qualifiers	/experiment="[CATEGORY:]text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /map="text" /note="text"
Comment	The location span of the gap feature for an unknown gap is 100 bp, with the 100 bp indicated as 100 "n"'s in the sequence. Where estimated length is indicated by an integer, this is indicated by the

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	same number of "n"'s in the sequence. No upper or lower limit is set on the size of the gap.
Feature Key	**gene**
Definition	Region of biological interest identified as a gene and for which a name has been assigned.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text" /product="text" /pseudo /pseudogene="TYPE" /phenotype="text" /standard_name="text" /trans_splicing
Comment	The gene feature describes the interval of DNA that corresponds to a genetic trait or phenotype; the feature is, by definition, not strictly bound to it's positions at the ends; it is meant to represent a region where the gene is located.
Feature Key	**iDNA**
Definition	Intervening DNA; DNA which is eliminated through any of several kinds of recombination.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text"

	/function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /number=unquoted text (single token) /old_locus_tag="text" (single token) /standard_name="text"
Molecule scope	DNA
Comment	e.g., in the somatic processing of immunoglobulin genes.
Feature Key	**intron**
Definition	A segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text” /note="text" /number=unquoted text (single token) /old_locus_tag="text" (single token) /pseudo /pseudogene="TYPE" /standard_name="text" /trans_splicing

Feature Key	**J_segment**
Definition	Joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Organism scope	eukaryotes
Feature Key	**mat_peptide**
Definition	Mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification; the location does not include the stop codon (unlike the corresponding CDS).
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /EC_number="text" /experiment="[CATEGORY:]text" /function="text" /gene="text” /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token)

	/map="text” /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Feature Key	**misc_binding**
Definition	Site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other binding key (primer_bind or protein_bind).
Mandatory qualifiers	/bound_moiety="text"
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token)
Comment	Note that feature key regulatory with /regulatory_class="ribosome_binding_site" should be used for ribosome binding sites.
Feature Key	**misc_difference**
Definition	Feature sequence is different from that presented in the entry and cannot be described by any other difference key (old_sequence, variation, or modified_base).
Optional qualifiers	/allele="text" /citation=[number] /clone="text"

	/compare=[accession-number.sequence-version] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /phenotype="text" /replace="text" /standard_name="text"
Comment	The misc_difference feature key should be used to describe variability that arises as a result of genetic manipulation (e.g. site directed mutagenesis); use /replace="" to annotate deletion, e.g. misc_difference 412..433 /replace="".
Feature Key	**misc_feature**
Definition	Region of biological interest which cannot be described by any other feature key; a new or rare feature.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /number=unquoted text (single token) /old_locus_tag="text" (single token) /phenotype="text"

	/product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Comment	This key should not be used when the need is merely to mark a region in order to comment on it or to use it in another feature's location.
Feature Key	**misc_recomb**
Definition	Site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys or qualifiers of source key (/proviral)
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /recombination_class="TYPE" /standard_name="text"
Molecule scope	DNA
Feature Key	**misc_RNA**
Definition	Any transcript or RNA product that cannot be defined by other RNA keys prim_transcript, precursor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, ncRNA, rRNA and tRNA).
Optional qualifiers	/allele="text"

	/citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text" /product="text" /pseudo /pseudogene="TYPE" /standard_name="text” /trans_splicing
Feature Key	**misc_structure**
Definition	Any secondary or tertiary nucleotide structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop).
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /standard_name="text"

Feature Key	**mobile_element**
Definition	Region of genome containing mobile elements
Mandatory qualifiers	/mobile_element_type="<mobile_element_type> [:<mobile_element_name>]"
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text” /note="text" /old_locus_tag="text" (single token) /rpt_family="text" /rpt_type=<repeat_type> /standard_name="text"
Feature Key	**modified_base**
Definition	The indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value).
Mandatory qualifiers	/mod_base=<modified_base>
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /frequency="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token)

	/map="text" /note="text" /old_locus_tag="text" (single token)
Comment	Value is limited to the restricted vocabulary for modified base abbreviations.
Feature Key	**mRNA**
Definition	Messenger RNA; includes 5'untranslated region (5'UTR), coding sequences (CDS, exon) and 3'untranslated región (3'UTR).
Optional qualifiers	/allele="text" /artificial_location="[artificial_location_value]" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text" /product="text" /pseudo /pseudogene="TYPE" /standard_name="text" /trans_splicing
Feature Key	**ncRNA**
Definition	A non-protein-coding gene, other than ribosomal RNA and transfer RNA, the functional molecule of which is the RNA transcript.
Mandatory qualifiers	/ncRNA_class="TYPE"
Optional qualifiers	/allele="text"

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	/citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text" /product="text" /pseudo /pseudogene="TYPE" /standard_name="text" /trans_splicing
Example	/ncRNA_class="miRNA" /ncRNA_class="siRNA" /ncRNA_class="scRNA"
Comment	The ncRNA feature is not used for ribosomal and transfer RNA annotation, for which the rRNA and tRNA feature keys should be used, respectively.
Feature Key	**N_region**
Definition	Extra nucleotides inserted between rearranged immunoglobulin segments
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token)

	/map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE” /standard_name="text"
Organism scope	eukaryotes
Feature Key	**old_sequence**
Definition	The presented sequence revises a previous version of the sequence at this location
Mandatory qualifiers	/citation=[number] or /compare=[accession-number.sequence-version]
Optional qualifiers	/allele="text" /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /replace="text"
Comment	/replace="" is used to annotate deletion, e.g. old_sequence 12..15 /replace="" NOTE: This feature key is not valid in entries/records created from 15-Oct-2007.
Feature Key	**operon**
Definition	Region containing polycistronic transcript including a cluster ofgenes that are under the control of the same regulatory sequences/promoter and in the same biological pathway.
Mandatory qualifiers	/operon="text"

Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /map="text" /note="text" /phenotype="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Feature Key	**oriT**
Definition	Origin of transfer; region of a DNA molecule where transfer is initiated during the process of conjugation or mobilization.
Optional qualifiers	/allele="text" /bound_moiety="text" /citation=[number] /db_xref="<database>:<identifier>" /direction=value /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /rpt_family="text” /rpt_type=<repeat_type> /rpt_unit_range=<base_range> /rpt_unit_seq="text"

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	/standard_name="text"
Molecule scope	DNA
Comment	rep_origin should be used for origins of replication; /direction has legal values RIGHT, LEFT and BOTH, however only RIGHT and LEFT are valid when used in conjunction with the oriT feature; origins of transfer can be present in the chromosome; plasmids can contain multiple origins of transfer.
Feature Key	**polyA_site**
Definition	Site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token)
Organism scope	eukaryotes and eukaryotic viruses
Feature Key	**precursor_RNA**
Definition	Any RNA species that is not yet the mature RNA product; may include ncRNA, rRNA, tRNA, 5' untranslated región (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR).
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text"

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	/gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text" /product="text" /standard_name="text" /trans_splicing
Comment	Used for RNA which may be the result of post-transcriptional processing; if the RNA in question is known not to have been processed, use the prim_transcript key.
Feature Key	**prim_transcript**
Definition	Primary (initial, unprocessed) transcript may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR).
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text” /standard_name="text"

Feature Key	**primer_bind**
Definition	Non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic e.g., PCR primer elements.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /standard_name="text" /PCR_conditions="text"
Comment	Used to annotate the site on a given sequence to which a primer molecule binds - not intended to represent the sequence of the primer molecule itself; PCR components and reaction times may be stored under the "/PCR_conditions" qualifier; since PCR reactions most often involve pairs of primers, a single primer_bind key may use the order() operator with two locations, or a pair of primer_bind keys may be used.
Feature Key	**propeptide**
Definition	Propeptide coding sequence; coding sequence for the domain of a proprotein that is cleaved to form the mature protein product.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text"

	/gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Feature Key	**protein_bind**
Definition	Non-covalent protein binding site on nucleic acid
Mandatory qualifiers	/bound_moiety="text"
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text" /standard_name="text"
Comment	Note that feature key regulatory with /regulatory_class="ribosome_binding_site" should be used for ribosome binding sites.
Feature Key	**regulatory**
Definition	Any region of sequence that functions in the regulation of transcription, translation, replication or chromatin structure.

Mandatory qualifiers	/regulatory_class="TYPE"
Optional qualifiers	/allele="text" /bound_moiety="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text" /phenotype="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Comment	This feature has replaced the following Feature Keys on 15-DEC-2014: enhancer, promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, RBS, GC_signal, polyA_signal, attenuator, terminator, misc_signal.
Feature Key	**repeat_region**
Definition	Region of genome containing repeating units.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]"

	/locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /rpt_family="text" /rpt_type=<repeat_type> /rpt_unit_range=<base_range> /rpt_unit_seq="text" /satellite="<satellite_type>[:<class>][<identifier>]" /standard_name="text"
Feature Key	**rep_origin**
Definition	Origin of replication; starting site for duplication of nucleic acid to give two identical copies.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /direction=value /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /standard_name="text"
Comment	/direction has valid values: RIGHT, LEFT, or BOTH.
Feature Key	**rRNA**
Definition	Mature ribosomal RNA; RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins.
Optional qualifiers	/allele="text” /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text”

	/function="text" /gene="text” /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /operon="text” /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Comment	rRNA sizes should be annotated with the /product qualifier.
Feature Key	**S_region**
Definition	Switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Organism scope	eukaryotes

Feature Key	**sig_peptide**
Definition	Signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane leader sequence.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Feature Key	**source**
Definition	Identifies the biological source of the specified span of the sequence; this key is mandatory; more than one source key per sequence is allowed; every entry/record will have, as a minimum, either a single source key spanning the entire sequence or multiple source keys, which together, span the entire sequence.
Mandatory qualifiers	/organism="text” /mol_type="genomic DNA", "genomic RNA", "mRNA", "tRNA", "rRNA", "other RNA", "other DNA", "transcribed RNA", "viral cRNA", "unassigned DNA", "unassigned RNA"
Optional qualifiers	/altitude="text" /bio_material="[<institution-code>:[<collection-code>:]]<material_id>" /cell_line="text" /cell_type="text"

	/chromosome="text" /citation=[number] /clone="text" /clone_lib="text" /collected_by="text" /collection_date="text" /country="<country_value>[:<region>][, <locality>]" /cultivar="text" /culture_collection="<institution-code>:[<collection-code>:]<culture_id>" /db_xref="<database>:<identifier>" /dev_stage="text" /ecotype="text" /environmental_sample /focus /germline /haplogroup="text" /haplotype="text” /host="text” /identified_by="text” /isolate="text” /isolation_source="text” /lab_host="text” /lat_lon="text" /macronuclear /map="text” /mating_type="text” /note="text” /organelle=<organelle_value> /PCR_primers="[fwd_name: XXX,]fwd_seq: xxxxx, [rev_name: YYY,]rev_seq: yyyyy” /plasmid="text” /pop_variant="text" /proviral /rearranged /segment="text” /serotype="text”

	/serovar="text" /sex="text” /specimen_voucher="[<institution-code>:[<collection-code>:]]<specimen_id>" /strain="text" /sub_clone="text" /submitter_seqid="text" /sub_species="text" /sub_strain="text” /tissue_lib="text" /tissue_type="text” /transgenic /type_material="<type-of-type> of <organism name>" /variety="text"
Molecule scope	any
Comment	Transgenic sequences must have at least two source feature keys; in a transgenic sequence the source feature key describing the organism that is the recipient of the DNA must span the entire sequence; see Appendix III /organelle for a list of <organelle_value>.
Feature Key	**stem_loop**
Definition	Hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA.
Optional qualifiers	/allele="text” /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /function="text" /gene="text” /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text”

	/old_locus_tag="text" (single token) /operon="text” /standard_name="text"
Feature Key	**STS**
Definition	Sequence tagged site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSs.
Optional qualifiers	/allele="text” /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /gene="text” /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /standard_name="text"
Molecule scope	DNA
Comment	STS location to include primer(s) in primer_bind key or primers.
Feature Key	**telomere**
Definition	Region of biological interest identified as a telomere and which has been experimentally characterized.
Optional qualifiers	/citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text"/note="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /note="text" /rpt_type=<repeat_type> /rpt_unit_range=<base_range> /rpt_unit_seq="text” /standard_name="text"

Comment	The telomere feature describes the interval of DNA that corresponds to a specific structure at the end of the linear eukaryotic chromosome which is required for the integrity and maintenance of the end; this región is unique compared to the rest of the chromosome and represent the physical end of the chromosome.
Feature Key	**tmRNA**
Definition	Transfer messenger RNA; tmRNA acts as a tRNA first, and then as an mRNA that encodes a peptide tag; the ribosome translates this mRNA region of tmRNA and attaches the encoded peptide tag to the C-terminus of the unfinished protein; this attached tag targets the protein for destruction or proteolysis.
Optional qualifiers	/allele="text” /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /function="text" /gene="text” /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE” /standard_name="text” /tag_peptide=<base_range>
Feature Key	**transit_peptide**
Definition	Transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle.

Optional qualifiers	/allele="text” /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /function="text" /gene="text” /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE” /standard_name="text"
Feature Key	**tRNA**
Definition	Mature transfer RNA, a small RNA molecule (75-85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence.
Optional qualifiers	/allele="text" /anticodon=(pos:<location>,aa:<amino_acid>,seq:<text>) /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /operon="text"

	/product="text" /pseudo /pseudogene="TYPE” /standard_name="text” /trans_splicing
Feature Key	**Unsure**
Definition	A small region of sequenced bases, generally 10 or fewer in its length, which could not be confidently identified. Such a region might contain called bases (A, T, G, or C), or a mixture of called-bases and uncalled-bases ('N'). The unsure feature should not be used when annotating gaps in genome assemblies. Please refer to assembly_gap feature for gaps within the sequence of an assembled genome. For annotation of gaps in other sequences than assembled genomes use the gap feature.
Optional qualifiers	/allele="text" /citation=[number] /compare=[accession-number.sequence-version] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /replace="text"
Comment	use /replace="" to annotate deletion, e.g unsure 11..15 /replace=""
Feature Key	**V_region**
Definition	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be composed of V_segments, D_segments, N_regions, and J_segments

Optional qualifiers	allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASICS]” /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /pseudogene="TYPE" /standard_name="text"
Molecule scope	eukaryotes
Feature Key	**V_segment**
Definition	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /product="text" /pseudo

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	/pseudogene="TYPE” /standard_name="text"
Molecule scope	eukaryotes
Feature Key	**variation**
Definition	A related strain contains stable mutations from the same gene (e.g., RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others).
Optional qualifiers	/allele="text" /citation=[number] /compare=[accession-number.sequence-version] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /frequency="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /phenotype="text" /product="text" /replace="text” /standard_name="text"
Comment	Used to describe alleles, RFLP's,and other naturally occurring mutations and polymorphisms; variability arising as a result of genetic manipulation (e.g. site directed mutagenesis) should be described with the misc_difference feature; use /replace="" to annotate deletion, e.g. variation 4..5 /replace="".
Feature Key	**3'UTR**
Definition	1) Region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein; 2) region at the 3' end of an RNA virus (following the last stop codon) that is not translated into a protein.

Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /standard_name="text” /trans_splicing
Feature Key	**5'UTR**
Definition	1) Region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein; 2) region at the 5' end of an RNA virus genome (preceding the first initiation codon) that is not translated into a protein.
Optional qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text” /function="text" /gene="text" /gene_synonym="text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /locus_tag="text" (single token) /map="text" /note="text” /old_locus_tag="text" (single token) /standard_name="text” /trans_splicing

LISTA DE "QUALIFIERS" Y SU DESCRIPCIÓN

La información obedece el siguiente formato:

Qualifier: nombre del calificador, que requiere de un valor si contiene un signo igual.

Definition: descripción del calificador.

Value format: formato del valor, si es necesario.

Example: ejemplo del calificador con valor.

Comment: comentarios y aclaraciones.

Qualifier	**/allele=**
Definition	Name of the allele for the given gene.
Value format	"text"
Example	/allele="adh1-1"
Comment	All gene-related features (exon, CDS etc) for a given gene should share the same /allele qualifier value; the /allele qualifier value must, by definition, be different from the /gene qualifier value; when used with the variation feature key, the allele qualifier value should be that of the variant.
Qualifier	**/altitude=**
Definition	Geographical altitude of the location from which the simple was collected.
Value format	"text"
Example	/altitude="-256 m" /altitude="330.12 m"
Comment	Values indicate altitudes above or below nominal sea level provided in metres.
Qualifier	**/anticodon=**
Definition	Location of the anticodon of tRNA and the amino acid for which it codes.
Value format	(pos:<location>,aa:<amino_acid>,seq:<text>) where location is the position of the anticodon and amino_acid is the abbreviation for the amino acid encoded and seq is the sequence of the anticodon.
Example	/anticodon=(pos:34..36,aa:Phe,seq:aaa) /anticodon=(pos:join(5,495..496),aa:Leu,seq:taa)

	/anticodon=(pos:complement(4156..4158),aa:Gln,seq:ttg)
Qualifier	**/artificial_location**
Definition	Indicates that location of the CDS or mRNA is modified to adjust for the presence of a frameshift or internal stop codon and not because of biological processing between the regions.
Value format	"heterogeneous population sequenced", "low-quality sequence region"
Example	/artificial_location="heterogeneous population sequenced" /artificial_location="low-quality sequence region"
Comment	Expected to be used only for genome-scale annotation.
Qualifier	**/bio_material=**
Definition	Dentifier for the biological material from which the nucleic acid sequenced was obtained, with optional institution code and collection code for the place where it is currently stored.
Value format	"[<institution-code>:[<collection-code>:]]<material_id>"
Example	**/bio_material="CGC:CB3912" <- Caenorhabditis stock centre**
Comment	The bio_material qualifier should be used to annotate the identifiers of material in biological collections that are not appropriate to annotate as either /specimen_voucher or /culture_collection; these include zoos and aquaria, stock centres, seed banks, germplasm repositories and DNA banks; material_id is mandatory, institution_code and collection_code are optional; institution code is mandatory where collection code is present; institution code and collection code are taken from a controlled vocabulary maintained by the INSDC.
Qualifier	**/bound_moiety=**
Definition	Name of the molecule/complex that may bind to the given feature.
Value format	"text"
Example	/bound_moiety="GAL4"
Comment	A single /bound_moiety qualifier is legal on the "misc_binding", "oriT" and "protein_bind" features.
Qualifier	**/cell_line=**
Definition	Cell line from which the sequence was obtained.
Value format	"text"

Example	/cell_line="MCF7"
Qualifier	**/cell_type=**
Definition	Cell type from which the sequence was obtained.
Value format	"text"
Example	/cell_type="leukocyte"
Qualifier	**/chromosome=**
Definition	Chromosome (e.g. Chromosome number) from whichthe sequence was obtained.
Value format	"text"
Example	/chromosome="1"
Qualifier	**/citation=**
Definition	Reference to a citation listed in the entry reference field.
Value format	[integer-number] where integer-number is the number of the reference as enumerated in the reference field.
Example	/citation=[3]
Comment	Used to indicate the citation providing the claim of and/or evidence for a feature; brackets are used for conformity.
Qualifier	**/clone=**
Definition	Clone from which the sequence was obtained.
Value format	"text"
Example	/clone="lambda-hIL7.3"
Comment	Not more than one clone should be specified for a given source feature; to indicate that the sequence was obtained from multiple clones, multiple source features should be given.
Qualifier	**/clone_lib=**
Definition	Clone library from which the sequence was obtained.
Value format	"text"
Example	/clone_lib="lambda-hIL7"
Qualifier	**/codon_start=**
Definition	Indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature.

Value format	1 or 2 or 3
Example	/codon_start=2
Qualifier	**/collected_by=**
Definition	Name of persons or institute who collected the specimen.
Value format	"text"
Example	/collected_by="Dan Janzen"
Qualifier	**/collection_date=**
Definition	The date on which the specimen was collected. Date/time ranges are supported by providing two collection dates from among the supported value formats, delimited by a forward-slash character. Collection times are supported by adding "T", then the hour and minute and seconds, after the date. Collection times must be in Coordinated Universal Time (UTC), otherwise known as "Zulu Time" (Z).
Value format	"DD-Mmm-YYYY", "Mmm-YYYY", "YYYY" "YYYY-MM-DDThh:mmZ", "YYYY-MM-DDThh:mm:ssZ", "YYYY-MM-DDThhZ", "YYYY-MM-DD", or "YYYY-MM"
Example	/collection_date="21-Oct-1952” /collection_date="Oct-1952” /collection_date="1952” /collection_date="1952-10-21T11:43Z” /collection_date="1952-10-21T11Z” /collection_date="1952-10-21” /collection_date="1952-10” /collection_date="21-Oct-1952/15-Feb-1953" /collection_date="Oct-1952/Feb-1953" /collection_date="1952/1953” /collection_date="1952-10-21/1953-02-15" /collection_date="1952-10/1953-02” /collection_date="1952-10-21T11:43Z/1952-10-21T17:43Z" /collection_date="2015-10-11T17:53:03Z"
Comment	'Mmm' represents a three-letter month abbreviation, and can be one of the following: Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec.

	'YYYY' is a four-digit value representing the year. 'MM' is a two-digit value representing the month. 'DD' is a two-digit value representing the day of the month. 'hh' is a two-digit value representing the hour of the day (00 to 23). 'mm' is a two-digit value representing the minute of the hour (00 to 59). 'ss' is a two-digit value representing the second of the hour (00 to 59). Within a date range, value formats that make use of 'Mmm' (month abbreviations) cannot be combined with value formats that make use of 'MM' (two-digit month number). Collection dates that are specified to at least the month, day, and year (DD-Mmm-YYYY or YYYY-MM-DD) are strongly encouraged. If the day and/or month of the collection date are not known, Mmm-YYYY or YYYY-MM or YYYY may be used. Within a collection date range, the first date (possibly including time) must be prior to the second date (possibly including time). Within a collection date range for which the day, month, and year are identical, the first time value must be prior to the second time value.
Qualifier	**/compare=**
Definition	Reference details of an existing public INSD entry to which a comparison is made.
Value format	[accession-number.sequence-version]
Example	/compare=AJ634337.1
Comment	This qualifier may be used on the following features: misc_difference, unsure, old_sequence and variation. The feature "old_sequence" must have either a /citation or a /compare qualifier. Multiple /compare qualifiers with different contents are allowed within a single feature. This qualifier is not intended for large-scale annotation of variations, such as SNPs.
Qualifier	**/country=**
Definition	Locality of isolation of the sequenced organism indicated in terms of political names for nations, oceans or seas, followed by regions and localitie.

Value format	"<country_value>[:<region>][, <locality>]" where country_value is any value from the controlled vocabulary at http://www.insdc.org/documents/country-qualifier-vocabulary
Example	/country="Canada:Vancouver" /country="France:Cote d'Azur, Antibes" /country="Atlantic Ocean:Charlie Gibbs Fracture Zone"
Comment	Intended to provide a reference to the site where the source organism was isolated or sampled. Regions and localities should be indicated where possible. Note that the physical geography of the isolation or sampling site should be represented in /isolation_source.
Qualifier	**/cultivar=**
Definition	Cultivar (cultivated variety) of plant from which sequence was obtained.
Value format	"text"
Example	/cultivar="Nipponbare" /cultivar="Tenuifolius" /cultivar="Candy Cane" /cultivar="IR36"
Comment	'cultivar' is applied solely to products of artificial selection; use the variety qualifier for natural, named plant and fungal varieties.
Qualifier	**/culture_collection=**
Definition	Institution code and identifier for the culture from which the nucleic acid sequenced was obtained, with optional collection code.
Value format	"<institution-code>:[<collection-code>:]<culture_id>"
Example	/culture_collection="ATCC:26370"
Comment	The /culture_collection qualifier should be used to annotate live microbial and viral cultures, and cell lines that have been deposited in curated culture collections; microbial cultures in personal or laboratory collections should be annotated in strain qualifiers; annotation with a culture_collection qualifier implies that the sequence was obtained from a sample retrieved (by the submitter or a collaborator) from the indicated culture collection, or that the sequence was obtained from a sample that was deposited (by the submitter or a collaborator) in the indicated culture collection; annotation with more than one culture_collection qualifier indicates that the sequence was obtained from a simple that was

	deposited (by the submitter or a collaborator) in more than one culture collection. culture_id and institution_code are mandatory, collection_code is optional; institution code and collection code are taken from a controlled vocabulary maintained by the INSDC. http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier
Qualifier	**/db_xref=**
Definition	Database cross-reference: pointer to related information in another database.
Value format	"<database:identifier>" where database is the name of the database containing related information, and identifier is the internal identifier of the related information according to the naming conventions of the cross-referenced database.
Example	/db_xref="UniProtKB/Swiss-Prot:P28763"
Comment	The complete list of allowed database types is kept at http://www.insdc.org/db_xref.html
Qualifier	**/dev_stage=**
Definition	If the sequence was obtained from an organism in a specific developmental stage, it is specified with this qualifier.
Value format	"text"
Example	/dev_stage="fourth instar larva"
Qualifier	**/direction=**
Definition	Direction of DNA replication.
Value format	Left, right, or both where left indicates toward the 5' end of the entry sequence (as presented) and right indicates toward the 3' end.
Example	/direction=LEFT
Qualifier	**/EC_number=**
Definition	Enzyme Commission number for enzyme product of sequence.
Value format	"text"
Example	/EC_number="1.1.2.4" /EC_number="1.1.2.-" /EC_number="1.1.2.n"

Comment	Valid values for EC numbers are defined in the list prepared by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (published in Enzyme Nomenclature 1992, Academic Press, San Diego, or a more recent revision thereof). The format represents a string of four numbers separated by full stops; up to three numbers starting from the end of the string can be replaced by dash "." to indicate uncertain assignment. Symbol "n" can be used in the last position instead of a number where the EC number is awaiting assignment. Please note that such incomplete EC numbers are not approved by NC-IUBMB.
Qualifier	**/ecotype=**
Definition	A population within a given species displaying genetically based, phenotypic traits that reflect adaptation to a local hábitat.
Value format	"text"
Example	/ecotype="Columbia"
Comment	An example of such a population is one that has adapted hairier than normal leaves as a response to an especially sunny habitat. 'Ecotype' is often applied to standard genetic stocks of Arabidopsis thaliana, but it can be applied to any sessile organism.
Qualifier	**/environmental_sample**
Definition	Identifies sequences derived by direct molecular isolation from a bulk environmental DNA sample (by PCR with or without subsequent cloning of the product, DGGE, or other anonymous methods) with no reliable identification of the source organism. Environmental samples include clinical samples, gut contents, and other sequences from anonymous organisms that may be associated with a particular host. They do not include endosymbionts that can be reliably recovered from a particular host, organisms from a readily identifiable but uncultured field sample (e.g., many cyanobacteria), or phytoplasmas that can be reliably recovered from diseased plants (even though these cannot be grown in axenic culture).
Value format	none
Example	/environmental_sample

Comment	Used only with the source feature key; source feature keys containing the /environmental_sample qualifier should also contain the /isolation_source qualifier. entries including /environmental_sample must not include the /strain qualifier.
Qualifier	**/estimated_length=**
Definition	Estimated length of the gap in the sequence.
Value format	unknown or <integer>
Example	/estimated_length=unknown /estimated_length=342
Qualifier	**/exception=**
Definition	Indicates that the coding region cannot be translated using standard biological rules.
Value format	"RNA editing", "reasons given in citation", "rearrangement required for product", "annotated by transcript or proteomic data"
Example	/exception="RNA editing” /exception="reasons given in citation" /exception="rearrangement required for product" /exception="annotated by transcript or proteomic data"
Comment	Only to be used to describe biological mechanisms such as RNA editing; where the exception cannot easily be described a published citation must be referred to; protein translation of /exception CDS will be different from the according conceptual translation; - An /inference qualifier should accompany any use of /exception="annotated by transcript or proteomic data", to provide support for the existence of the transcript/protein. - Must not be used where transl_except would be adequate, e.g. in case of stop codon completion use: /transl_except=(pos:6883,aa:TERM) /note="TAA stop codon is completed by addition of 3' A residues to mRNA".

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	- Must not be used for ribosomal slippage, instead use join operator, e.g.: CDS join(486..1784,1787..4810) /note="ribosomal slip on tttt sequence at 1784..1787".
Qualifier	**/experiment=**
Definition	A brief description of the nature of the experimental evidence that supports the feature identification or assignment.
Value format	"[CATEGORY:]text" Where CATEGORY is one of the following: "COORDINATES" support for the annotated coordinates "DESCRIPTION" support for a broad concept of function such as that based on phenotype, genetic approach, biochemical function, pathway information, etc. "EXISTENCE" support for the known or inferred existence of the product where text is free text (see examples).
Example	/experiment="5' RACE” /experiment="Northern blot [DOI: 12.3456/FT.789.1.234-567.2010]" /experiment="heterologous expression system of Xenopus laevis oocytes [PMID: 12345678, 10101010, 987654]" /experiment="COORDINATES: 5' and 3' RACE"
Comment	Detailed experimental details should not be included, and would normally be found in the cited publications; PMID, DOI and any experimental database ID is allowed to be used in /experiment qualifier; value "experimental evidence, no additional details recorded" was used to replace instances of /evidence=EXPERIMENTAL in December 2005.
Qualifier	**/focus**
Definition	Identifies the source feature of primary biological interest for records that have multiple source features originating from different organisms and that are not transgenic.
Value format	none
Example	/focus

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Comment	The source feature carrying the /focus qualifier identifies the main organism of the entry, this determines: a) the name displayed in the organism lines, b) if no translation table is specified, the translation table, c) the DDBJ/EMBL/GenBank taxonomic division in which the entry will appear; only one source feature with /focus is allowed in an entry; the /focus and /transgenic qualifiers are mutually exclusive in an entry.
Qualifier	**/frequency=**
Definition	Frequency of the occurrence of a feature.
Value format	Text representing the proportion of a population carrying the feature expressed as a fraction.
Example	/frequency="23/108" /frequency="1 in 12" /frequency=".85"
Qualifier	**/function=**
Definition	Function attributed to a sequence.
Value format	"text"
Example	function="essential for recognition of cofactor"
Comment	/function is used when the gene name and/or product name do not convey the function attributable to a sequence.
Qualifier	**/gap_type=**
Definition	Type of gap connecting components in records of a genome assembly, or the type of biological gap in a record that is part of a genome assembly.
Value format	"between scaffolds", "within scaffold", "telomere", "centromere", "short arm", "heterochromatin", "repeat within scaffold", "repeat between scaffolds", "contamination", "unknown"
Example	/gap_type="between scaffolds” /gap_type="within scaffold"
Comment	This qualifier is used only for assembly_gap features and its values are controlled by the AGP Specification version 2.0: https://www.ncbi.nlm.nih.gov/assembly/agp/AGP_Specification/ Please also visit: http://www.insdc.org/controlled-vocabulary-gaptype-qualifier
Qualifier	**/gene=**
Definition	Symbol of the gene corresponding to a sequence región.

Value format	"text"
Example	/gene="ilvE"
Qualifier	**/gene_synonym=**
Definition	Synonymous, replaced, obsolete or former gene symbol.
Value format	"text"
Example	/gene_synonym="Hox-3.3" in a feature where /gene="Hoxc6"
Comment	Used where it is helpful to indicate a gene symbol synonym; when used, a primary gene symbol must always be indicated in /gene or a /locus_tag must be used.
Qualifier	**/germline**
Definition	The sequence presented in the entry has not undergone somatic rearrangement as part of an adaptive immune response; it is the unrearranged sequence that was inherited from the parental germline.
Value format	none
Example	/germline
Comment	/germline should not be used to indicate that the source of the sequence is a gamete or germ cell; /germline and /rearranged cannot be used in the same source feature; /germline and /rearranged should only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); /germline and /rearranged should not be used outside of the Craniata (taxid=89593).
Qualifier	**/haplogroup=**
Definition	Name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups.
Value format	"text"
Example	/haplogroup="H*"

Qualifier	**/haplotype=**
Definition	Name for a combination of alleles that are linked together on the same physical chromosome. In the absence of recombination, each haplotype is inherited as a unit, and may be used to track gene flow in populations.
Value format	"text"
Example	/haplotype="Dw3 B5 Cw1 A1"
Qualifier	**/host=**
Definition	Natural (as opposed to laboratory) host to the organism from which sequenced molecule was obtained.
Value format	"text"
Example	/host="Homo sapiens" /host="Homo sapiens 12 year old girl” /host="Rhizobium NGR234"
Qualifier	**/identified_by=**
Definition	Name of the expert who identified the specimen taxonomically.
Value format	"text"
Example	/identified_by="John Burns"
Qualifier	**/inference=**
Definition	A structured description of non-experimental evidence that supports the feature identification or assignment.
Value format	"[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" where CATEGORY is one of the following: "COORDINATES" support for the annotated coordinates "DESCRIPTION" support for a broad concept of function such as that based on phenotype, genetic approach, biochemical function, pathway information, etc. "EXISTENCE" support for the known or inferred existence of the product

	where TYPE is one of the following: "non-experimental evidence, no additional details recorded" "similar to sequence" "similar to AA sequence" "similar to DNA sequence" "similar to RNA sequence" "similar to RNA sequence, mRNA" "similar to RNA sequence, EST" "similar to RNA sequence, other RNA" "profile" "nucleotide motif" "protein motif" "ab initio prediction" "alignment" where the optional text "(same species)" is included when the inference comes from the same species as the entry. where the optional "EVIDENCE_BASIS" is either a reference to a database entry (including accession and version) or an algorithm (including version) , eg 'INSD:AACN010222672.1', 'InterPro:IPR001900', 'ProDom:PD000600', 'Genscan:2.0', etc. and is structured "[ALGORITHM][:EVIDENCE_DBREF[,EVIDENCE_DBREF]*[,...]]"
Example	/inference="COORDINATES:profile:tRNAscan:2.1" /inference="similar to DNA sequence:INSD:AY411252.1" /inference="similar to RNA sequence, mRNA:RefSeq:NM_000041.2" /inference="similar to DNA sequence (same species):INSD:AACN010222672.1” /inference="protein motif:InterPro:IPR001900” /inference="ab initio prediction:Genscan:2.0" /inference="alignment:Splign:1.0" /inference="alignment:Splign:1.26p:RefSeq:NM_000041.2,INSD:BC003557.1"
Comment	/inference="non-experimental evidence, no additional details recorded" was used to replace instances of /evidence=NOT_EXPERIMENTAL in

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	December 2005; any database ID can be used in /inference= qualifier; recommentations for choice of resource acronym for[EVIDENCE_BASIS] are provided in the /inference qualifier vocabulary recommendation document (http://www.insdc.org/inference.html).
Qualifier	**/isolate=**
Definition	Individual isolate from which the sequence was obtained.
Value format	"text"
Example	/isolate="Patient #152" /isolate="DGGE band PSBAC-13"
Qualifier	**/isolation_source=**
Definition	Describes the physical, environmental and/or local geographical source of the biological sample from which the sequence was derived.
Value format	"text"
Example	/isolation_source="rumen isolates from standard Pelleted ration-fed steer #67" /isolation_source="permanent Antarctic sea ice" /isolation_source="denitrifying activated sludge from carbon_limited continuous reactor"
Comment	Used only with the source feature key; source feature keys containing an /environmental_sample qualifier should also contain an /isolation_source qualifier; the /country qualifier should be used to describe the country and major geographical sub-region.
Qualifier	**/lab_host=**
Definition	Scientific name of the laboratory host used to propagate the source organism from which the sequenced molecule was obtained.
Value format	"text"
Example	/lab_host="Gallus gallus" /lab_host="Gallus gallus embryo” /lab_host="Escherichia coli strain DH5 alpha” /lab_host="Homo sapiens HeLa cells"
Comment	The full binomial scientific name of the host organism should be used when known; extra conditional information relating to the host may also be included.

Qualifier	**/lat_lon=**
Definition	Geographical coordinates of the location where the specimen was collected.
Value format	"text"
Example	/lat_lon="47.94 N 28.12 W" /lat_lon="45.0123 S 4.1234 E"
Comment	Degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E" (see the examples).
Qualifier	**/linkage_evidence=**
Definition	Type of evidence establishing linkage across an assembly_gap. Only allowed to be used with assembly_gap features that have a /gap_type value of "within scaffold"or "repeat within scaffold".
Value format	"pcr", "paired-ends", "align genus", "align xgenus", "align trnscpt", "within clone", "clone contig", "map", "strobe", "unspecified"
Example	/linkage_evidence="paired-ends" /linkage_evidence="within clone"
Comment	This qualifier is used only for assembly_gap features and its values are controlled by the AGP Specification version 2.0: https://www.ncbi.nlm.nih.gov/assembly/agp/AGP_Specification/ Please also visit: http://www.insdc.org/controlled-vocabulary-linkageevidence-qualifier
Qualifier	**/locus_tag=**
Definition	A submitter-supplied, systematic, stable identifier for a gene and its associated features, used for tracking purposes.
Value format	"text"(single token) but not "<1-5 letters><5-9 digit integer>[.<integer>]"
Example	/locus_tag="ABC_0022" /locus_tag="A1C_00001"
Comment	/locus_tag can be used with any feature that /gene can be used with; identical /locus_tag values may be used within an entry/record, but only if the identical /locus_tag values are associated with the same gene; in all other circumstances the /locus_tag value must be unique within that

	entry/record. Multiple /locus_tag values are not allowed within one feature for entries created after 15-OCT-2004. If a /locus_tag needs to be re-assigned the /old_locus_tag qualifier should be used to store the old value. The /locus_tag value should not be in a format which resembles INSD accession numbers, accession.version, or /protein_id identifiers.
Qualifier	**/macronuclear**
Definition	If the sequence shown is DNA and from an organism which undergoes chromosomal differentiation between macronuclear and micronuclear stages, this qualifier is used to denote that the sequence is from macronuclear DNA.
Value format	none
Example	/macronuclear
Qualifier	**/map=**
Definition	Genomic map position of feature.
Value format	"text"
Example	/map="8q12-q13"
Qualifier	**/mating_type=**
Definition	Mating type of the organism from which the sequence was obtained; mating type is used for prokaryotes, and for eukaryotes that undergo meiosis without sexually dimorphic gametes.
Value format	"text"
Example	/mating_type="MAT-1" /mating_type="plus" /mating_type="-" /mating_type="odd" /mating_type="even"
Comment	/mating_type="male" and /mating_type="female" are valid in the prokaryotes, but not in the eukaryotes; for more information, see the entry for /sex.

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Qualifier	**/mobile_element_type=**
Definition	Type and name or identifier of the mobile element which is described by the parent feature.
Value format	"<mobile_element_type>[:<mobile_element_name>]" where mobile_element_type is one of the following: "transposon", "retrotransposon", "integron", "insertion sequence", "non-LTR retrotransposon", "SINE", "MITE", "LINE", "other".
Example	/mobile_element_type="transposon:Tnp9"
Comment	/mobile_element_type is legal on mobile_element feature key only. Mobile element should be used to represent both elements which are currently mobile, and those which were mobile in the past. Value "other" requires a mobile_element_name.
Qualifier	**/mod_base=**
Definition	Abbreviation for a modified nucleotide base.
Value format	modified_base
Example	/mod_base=m5c
Comment	Modified nucleotides not found in the restricted vocabulary list can be annotated by entering '/mod_base=OTHER' with '/note="name of modified base"'.
Qualifier	**/mol_type=**
Definition	In vivo molecule type of sequence.
Value format	"genomic DNA", "genomic RNA", "mRNA", "tRNA", "rRNA", "other RNA", "other DNA", "transcribed RNA", "viral cRNA", "unassigned DNA", "unassigned RNA"
Example	/mol_type="genomic DNA"
Comment	All values refer to the in vivo or synthetic molecule for primary entries and the hypothetical molecule in Third PartyAnnotation entries; the value "genomic DNA" does not imply thatthe molecule is nuclear (e.g. organelle and plasmid DNA shouldbe described using "genomic DNA"); ribosomal RNA genes should be described using "genomic DNA"; "rRNA" should only be used if the ribosomal RNA molecule itself has been sequenced; /mol_type is mandatory on every source feature key; all /mol_type values within one entry/record must be the same; values "other RNA" and "other DNA" should be applied to synthetic molecules, values "unassigned

	DNA", "unassigned RNA" should be applied where in vivo molecule is unknown. Please also visit: http://www.insdc.org/controlled-vocabulary-moltype-qualifier
Qualifier	**/ncRNA_class=**
Definition	A structured description of the classification of the non-coding RNA described by the ncRNA parent key.
Value format	"TYPE"
Example	/ncRNA_class="miRNA" /ncRNA_class="siRNA" /ncRNA_class="scRNA"
Comment	TYPE is a term taken from the INSDC controlled vocabulary for ncRNA classes. For a complete list of supported values, please see: http://www.insdc.org/documents/ncrna-vocabulary; ncRNA classes not yet in the INSDC /ncRNA_class controlled vocabulary can be annotated by entering '/ncRNA_class="other"' with either '/product="[name of the product]"' OR '/note="[brief explanation of novel ncRNA_class]"'.
Qualifier	**/note=**
Definition	Any comment or additional information.
Value format	"text"
Example	/note="This qualifier is equivalent to a comment."
Qualifier	**/number=**
Definition	A number to indicate the order of genetic elements (e.g., exons or introns) in the 5' to 3' direction.
Value format	unquoted text (single token)
Example	/number=4 /number=6B

Comment	Text limited to integers, letters or combination of integers and/or letters represented as an unquoted single token (e.g. 5a, XIIb); any additional terms should be included in /standard_name. Example: /number=2ª /standard_name="long"
Qualifier	**/old_locus_tag=**
Definition	Feature tag assigned for tracking purposes.
Value format	"text" (single token)
Example	/old_locus_tag="RSc0382" /locus_tag="YPO0002"0
Comment	/old_locus_tag can be used with any feature where /gene is valid and where a /locus_tag qualifier is present. Identical /old_locus_tag values may be used within an entry/record, but only if the identical /old_locus_tag values are associated with the same gene; in all other circumstances the /old_locus_tag value must be unique within that entry/record. Multiple/old_locus_tag qualifiers with distinct values are allowed within a single feature; /old_locus_tag and /locus_tag values must not be identical within a single feature.
Qualifier	**/operon=**
Definition	Name of the group of contiguous genes transcribed into a single transcript to which that feature belongs.
Value format	"text"
Example	/operon="lac"
Qualifier	**/organelle=**
Definition	Type of membrane-bound intracellular structure from which the sequence was obtained.
Value format	chromatophore, hydrogenosome, mitochondrion, nucleomorph, plastid, mitochondrion:kinetoplast, plastid:chloroplast, plastid:apicoplast, plastid:chromoplast, plastid:cyanelle, plastid:leucoplast, plastid:proplastid
Example	/organelle="chromatophore” /organelle="hydrogenosome” /organelle="mitochondrion” /organelle="nucleomorph” /organelle="plastid”

	/organelle="mitochondrion:kinetoplast" /organelle="plastid:chloroplast” /organelle="plastid:apicoplast” /organelle="plastid:chromoplast” /organelle="plastid:cyanelle” /organelle="plastid:leucoplast” /organelle="plastid:proplastid"
Comment	Modifier text limited to values from controlled list. Please also visit: http://www.insdc.org/controlled-vocabulary-organelle-qualifier
Qualifier	**/organism=**
Definition	Scientific name of the organism that provided the sequenced genetic material.
Value format	"text"
Example	/organism="Homo sapiens"
Comment	The organism name which appears on the OS or ORGANISM line will match the value of the /organism qualifier of the source key in the simplest case of a one-source sequence.
Qualifier	**/partial**
Definition	Differentiates between complete regions and partial ones.
Value format	none
Example	/partial
Comment	Not to be used for new entries from 15-DEC-2001; use '<' and '>' signs in the location descriptors to indicate that the sequence is partial.
Qualifier	**/PCR_conditions=**
Definition	Description of reaction conditions and components for PCR.
Value format	"text"
Example	/PCR_conditions="Initial denaturation:94degC,1.5min"
Comment	Used with primer_bind key.
Qualifier	**/PCR_primers=**
Definition	PCR primers that were used to amplify the sequence. A single /PCR_primers qualifier should contain all the primers used for a single PCR reaction. If multiple forward or reverse primers are present in a

	single PCR reaction, multiple sets of fwd_name/fwd_seq or rev_name/rev_seq values will be present.
Value format	/PCR_primers="[fwd_name: XXX1,]fwd_seq: xxxxx1,[fwd_name: XXX2,]fwd_seq: xxxxx2, [rev_name: YYY1,]rev_seq: yyyyy1, [rev_name: YYY2,]rev_seq: yyyyy2"
Example	/PCR_primers="fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg" /PCR_primers=" fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i>gtgtatcttact" /PCR_primers="fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg"
Comment	fwd_seq and rev_seq are both mandatory; fwd_name and rev_name are both optional. Both sequences should be presented in 5'>3' order. The sequences should be given in the IUPAC degenerate-base alphabet, except for the modified bases; those must be enclosed within angle brackets <>.
Qualifier	**/phenotype=**
Definition	Phenotype conferred by the feature, where phenotype is defined as a physical, biochemical or behavioural characteristic or set of characteristics.
Value format	"text"
Example	/phenotype="erythromycin resistance"
Qualifier	**/plasmid=**
Definition	Name of naturally occurring plasmid from which the sequence was obtained, where plasmid is defined as an independently replicating genetic unit that cannot be described by /chromosome or /segment.
Value format	"text"
Example	/plasmid="C-589"
Qualifier	**/pop_variant=**
Definition	Name of subpopulation or phenotype of the sample from which the sequence was derived.
Value format	"text"

Example	/pop_variant="pop1" /pop_variant="Bear Paw"
Qualifier	**/product=**
Definition	Name of the product associated with the feature, e.g. the mRNA of an mRNA feature, the polypeptide of a CDS, the mature peptide of a mat_peptide, etc.
Value format	"text"
Example	/product="trypsinogen" (when qualifier appears in CDS feature) /product="trypsin" (when qualifier appears in mat_peptide feature) /product="XYZ neural-specific transcript" (when qualifier appears in mRNA feature)
Qualifier	**/protein_id=**
Definition	Protein identifier, issued by International collaborators. This qualifier consists of a stable ID portion (3+5 format with 3 position letters and 5 numbers) plus a version number after the decimal point.
Value format	<identifier>
Example	/protein_id="AAA12345.1"
Comment	When the protein sequence encoded by the CDS changes, only the version number of the /protein_id value is incremented; the stable part of the /protein_id remains unchanged and as a result will permanently be associated with a given protein; this qualifier is valid only on CDS features which translate into a valid protein.
Qualifier	**/proviral**
Definition	This qualifier is used to flag sequence obtained from a virus or phage that is integrated into the genome of another organism.
Value format	none
Example	/proviral
Qualifier	**/pseudo**
Definition	Indicates that this feature is a non-functional version of the element named by the feature key.
Value format	none
Example	/pseudo

Comment	The qualifier /pseudo should be used to describe non-functional genes that are not formally described as pseudogenes, e.g. CDS has no translation due to other reasons than pseudogenisation events. Other reasons may include sequencing or assembly errors. In order to annotate pseudogenes the qualifier /pseudogene= must be used indicating the TYPE which can be taken from the INSDC controlled vocabulary for pseudogenes.
Qualifier	**/pseudogene=**
Definition	Indicates that this feature is a pseudogene of the element named by the feature key.
Value format	"TYPE" where TYPE is one of the following: processed, unprocessed, unitary, allelic, unknown
Example	/pseudogene="processed” /pseudogene="unprocessed” /pseudogene="unitary" /pseudogene="allelic" /pseudogene="unknown"
Comment	TYPE is a term taken from the INSDC controlled vocabulary for pseudogenes (http://www.insdc.org/documents/pseudogene-qualifier-vocabulary): processed: the pseudogene has arisen by reverse transcription of a mRNA into cDNA, followed by reintegration into the genome. Therefore, it has lost any intron/exon structure, and it might have a pseudo-polyA-tail. unprocessed: the pseudogene has arisen from a copy of the parent gene by duplication followed by accumulation of random mutations. The changes, compared to their functional homolog, include insertions, deletions, premature stop codons, frameshifts and a higher proportion of non-synonymous versus synonymous substitutions. unitary: the pseudogene has no parent. It is the original gene, which is functional is some species but disrupted in some way (indels, mutation, recombination) in another species or strain.

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	allelic: a (unitary) pseudogene that is stable in the population but importantly it has a functional alternative allele also in the population. i.e., one strain may have the gene, another strain may have the pseudogene. MHC haplotypes have allelic pseudogenes. unknown: the submitter does not know the method of pseudogenisation.
Qualifier	**/rearranged**
Definition	The sequence presented in the entry has undergone somatic rearrangement as part of an adaptive immune response; it is not the unrearranged sequence that was inherited from the parental germline.
Value format	none
Example	/rearranged
Comment	/rearranged should not be used to annotate chromosome rearrangements that are not involved in an adaptive immune response; /germline and /rearranged cannot be used in the same source feature; /germline and /rearranged should only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); /germline and /rearranged should not be used outside of the Craniata (taxid=89593).
Qualifier	**/recombination_class**
Definition	A structured description of the classification of recombination hotspot region within a sequence.
Value format	"TYPE"
Example	/recombination_class="meiotic" /recombination_class="chromosome_breakpoint"

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Comment	TYPE is a term taken from the INSDC controlled vocabulary for recombination clases (http://www.insdc.org/controlled-vocabulary-recombination-class); in DEC 2017, the following terms were valid: "meiotic" "mitotic" "non_allelic_homologous” "chromosome_breakpoint" "other" Recombination classes not yet in the INSDC /recombination_class controlled vocabulary can be annotated by entering /recombination_class="other" with /note="[brief explanation of novel /recombination_class]".
Qualifier	**/regulatory_class**
Definition	A structured description of the classification of transcriptional, translational, replicational and chromatin structure related regulatory elements in a sequence.
Value format	"TYPE"
Example	/regulatory_class="promoter” /regulatory_class="enhancer" /regulatory_class="ribosome_binding_site"
Comment	TYPE is a term taken from the INSDC controlled vocabulary for regulatory classes. For a complete list of supported values, please see: http://www.insdc.org/controlled-vocabulary-regulatoryclass; regulatory classes not yet in the INSDC /regulatory_class controlled vocabulary can be annotated by entering /regulatory_class="other" with /note="[brief explanation of novel regulatory_class]".
Qualifier	**/replace=**
Definition	Indicates that the sequence identified a feature's intervals is replaced by the sequence shown in "text"; if no sequence is contained within the qualifier, this indicates a deletion.
Value format	"text"
Example	/replace="a"

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	/replace=""
Qualifier	**/ribosomal_slippage**
Definition	During protein translation, certain sequences can program ribosomes to change to an alternative reading frame by a mechanism known as ribosomal slippage.
Value format	none
Example	/ribosomal_slippage
Comment	A join operator, e.g.: [join(486..1784,1787..4810)] should be used in the CDS spans to indicate the location of ribosomal_slippage.
Qualifier	**/rpt_family=**
Definition	Type of repeated sequence; "Alu" or "Kpn", for example.
Value format	"text"
Example	/rpt_family="Alu"
Qualifier	**/rpt_type=**
Definition	Structure and distribution of repeated sequence.
Value format	tandem, direct, inverted, flanking, nested, dispersed, terminal, long_terminal_repeat, non_ltr_retrotransposon_polymeric_tract, centromeric_repeat, telomeric_repeat, x_element_combinatorial_repeat, y_prime_element and other.
Example	/rpt_type=INVERTED
Comment	The values are case-insensitive, i.e. both "INVERTED" and "inverted" are valid; For the most current list of allowed values and their definitions please visit: http://www.insdc.org/controlled-vocabulary-rpttype-qualifier
Qualifier	**/rpt_unit_range=**
Definition	Identity of a repeat range.
Value format	<base_range>
Example	/rpt_unit_range=202..245
Comment	Used to indicate the base range of the sequence that constitutes a repeated sequence specified by the feature keys oriT and repeat_region; qualifiers /rpt_unit_range and /rpt_unit_seq replaced qualifier /rpt_unit in December 2005.
Qualifier	**/rpt_unit_seq=**
Definition	Identity of a repeat sequence.

Value format	"text"
Example	/rpt_unit_seq="aagggc” /rpt_unit_seq="ag(5)tg(8)" /rpt_unit_seq="(AAAGA)6(AAAA)1(AAAGA)12"
Comment	Used to indicate the literal sequence that constitutes a repeated sequence specified by the feature keys oriT and repeat_region; qualifiers /rpt_unit_range and /rpt_unit_seq replaced qualifier /rpt_unit in December 2005.
Qualifier	**/satellite=**
Definition	Identifier for a satellite DNA marker, compose of many tándem repeats (identical or related) of a short basic repeated unit.
Value format	"<satellite_type>[:<class>][<identifier>]" where satellite_type is one of the following: "satellite", "microsatellite", "minisatellite"
Example	/satellite="satellite: S1a” /satellite="satellite: alpha” /satellite="satellite: gamma III” /satellite="microsatellite: DC130"
Comment	Many satellites have base composition or other properties that differ from those of the rest of the genome that allows them to be identified. Please also visit: http://www.insdc.org/controlled-vocabulary-satellite-qualifier
Qualifier	**/segment=**
Definition	Name of viral or phage segment sequenced.
Value format	"text"
Example	/segment="6"
Qualifier	**/serotype=**
Definition	Serological variety of a species characterized by its antigenic properties.
Value format	"text"
Example	/serotype="B1"
Comment	Used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for the prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Qualifier	**/serovar=**
Definition	Serological variety of a species (usually a prokaryote) characterized by its antigenic properties.
Value format	"text"
Example	/serovar="O157:H7"
Comment	Used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".
Qualifier	**/sex=**
Definition	Sex of the organism from which the sequence was obtained; sex is used for eukaryotic organisms that undergo meiosis and have sexually dimorphic gametes.
Value format	"text"
Example	sex="female" /sex="male" /sex="hermaphrodite" /sex="unisexual" /sex="bisexual" /sex="asexual" /sex="monoecious" [or monecious] /sex="dioecious" [or diecious]
Comment	/sex should be used (instead of /mating_type) in the Metazoa, Embryophyta, Rhodophyta & Phaeophyceae; /mating_type should be used (instead of /sex) in the Bacteria, Archaea & Fungi; neither /sex nor /mating_type should be used in the viruses; outside of the taxa listed above, /mating_type should be used unless the value of the qualifier is taken from the vocabulary given in the examples above.
Qualifier	**/specimen_voucher=**
Definition	Identifier for the specimen from which the nucleic acid sequenced was obtained.
Value format	/specimen_voucher="[<institution-code>:[<collection-code>:]]<specimen_id>"

Example	/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"
Comment	The /specimen_voucher qualifier is intended to annotate a reference to the physical specimen that remains after the sequence has been obtained; if the specimen was destroyed in the process of sequencing, electronic images (e-vouchers) are an adequate substitute for a physical voucher specimen; ideally the specimens will be deposited in a curated museum, herbarium, or frozen tissue collection, but often they will remain in a personal or laboratory collection for some time before they are deposited in a curated collection; there are three forms of specimen_voucher qualifiers; if the text of the qualifier includes one or more colons it is a 'structured voucher'; structured vouchers include institution-codes (and optional collection-codes) taken from a controlled vocabulary maintained by the INSDC that denotes the museum or herbarium collection where the specimen resides; Please also visit: http://www.insdc.org/controlled-vocabulary-specimenvoucher-qualifier
Qualifier	**/standard_name=**
Definition	Accepted standard name for this feature.
Value format	"text"
Example	/standard_name="dotted"
Comment	Use /standard_name to give full gene name, but use /gene to give gene symbol (in the above example /gene="Dt").
Qualifier	**/strain=**
Definition	Strain from which sequence was obtained.
Value format	"text"
Example	/strain="BALB/c"
Comment	Entries including /strain must not include the /environmental_sample qualifier.

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Qualifier	**/sub_clone=**
Definition	sub-clone from which sequence was obtained.
Value format	"text"
Example	/sub_clone="lambda-hIL7.20g"
Comment	The comments on /clone apply to /sub_clone.
Qualifier	**/submitter_seqid=**
Definition	Identifier attributed to each sequence within an assembly. This identifier is appropriate for WGS, TSA, TLS and CON records. The submitter_seqid should be unique within the context of a single set of assembled sequences.
Value format	"text"
Example	/submitter_seqid="NODE_1"
Comment	The length of the value should be limited to <51 charaters. Spaces, greater than (>), left/right square brackets ([ ]) and vertical bar (|) in addition to double quotation marks (") can not be used for the value of /submitter_seqid qualifier.
Qualifier	**/sub_species=**
Definition	Name of sub-species of organism from which sequence was obtained.
Value format	"text"
Example	/sub_species="lactis"
Qualifier	**/sub_strain=**
Definition	Name or identifier of a genetically or otherwise modified strain from which sequence was obtained, derived from a parental strain (which should be annotated in the /strain qualifier).sub_strain from which sequence was obtained.
Value format	"text"
Example	/sub_strain="abis"
Comment	If the parental strain is not given, this should be annotated in the strain qualifier instead of sub_strain. Either: /strain="K-12"

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

	/sub_strain="MG1655" or: /strain="MG1655"
Qualifier	**/tag_peptide=**
Definition	Base location encoding the polypeptide for proteolysis tag of tmRNA and its termination codon.
Value format	<base_range>
Example	/tag_peptide=90..122
Comment	It is recommended that the amino acid sequence corresponding to the /tag_peptide be annotated by describing a 5' partial CDS feature; e.g. CDS <90..122.
Qualifier	**/tissue_lib=**
Definition	Tissue library from which sequence was obtained.
Value format	"text"
Example	/tissue_lib="tissue library 772"
Qualifier	**/tissue_type=**
Definition	Tissue type from which the sequence was obtained.
Value format	"text"
Example	/tissue_type="liver"
Qualifier	**/transgenic**
Definition	Identifies the source feature of the organism which was the recipient of transgenic DNA.
Value format	none
Example	/transgenic
Comment	Transgenic sequences must have at least two source feature keys; the source feature key having the /transgenic qualifier must span the whole sequence; the source feature carrying the /transgenic qualifier identifies the main organism of the entry, this determines: a) the name displayed in the organism lines, b) if no translation table is specified, the translation table; only one source feature with /transgenic is allowed in an entry; the /focus and /transgenic qualifiers are mutually exclusive in an entry.

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Qualifier	**/translation=**
Definition	Automatically generated one-letter abbreviated amino acid sequence derived from either the universal genetic code or the table as specified in /transl_table and as determined by an exception in the /transl_except qualifier.
Value format	IUPAC one-letter amino acid abbreviation, "X" is to be used for AA exceptions.
Example	/translation="MASTFPPWYRGCASTPSLKGLIMCTW"
Comment	To be used with CDS feature only; this is a mandatory qualifier in the CDS feature key except where /pseudogene="TYPE" or /pseudo is shown; see /transl_table for definition and location of genetic code tables.
Qualifier	**/transl_except=**
Definition	Translational exception: single codon the translation of which does not conform to genetic code defined by /organism or /transl_table.
Value format	(pos:location,aa:<amino_acid>) where amino_acid is the amino acid coded by the codon at the base_range position.
Example	/transl_except=(pos:213..215,aa:Trp) /transl_except=(pos:1017,aa:TERM) /transl_except=(pos:2000..2001,aa:TERM) /transl_except=(pos:X22222:15..17,aa:Ala)
Comment	If the amino acid is not on the restricted vocabulary list use e.g., '/transl_except=(pos:213..215,aa:OTHER)' with '/note="name of unusual amino acid"'; for modified amino-acid selenocysteine use three letter code 'Sec' (one letter code 'U' in amino-acid sequence) /transl_except=(pos:1002..1004,aa:Sec); for partial termination codons where TAA stop codon is completed by the addition of 3' A residues to the mRNA either a single base_position or a base_range is used, e.g. if partial stop codon is a single base: /transl_except=(pos:1017,aa:TERM) if partial stop codon consists of two bases: /transl_except=(pos:2000..2001,aa:TERM) with '/note='stop codon completed by the addition of 3' A residues to the mRNA'.
Qualifier	**/transl_table=**
Definition	Definition of genetic code table used if other than universal genetic code table. Tables used are described in appendix IV.
Value format	<integer; 1=universal table 1;2=non-universal table 2;...

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Example	/transl_table=4
Comment	Genetic code exceptions outside range of specified tables are reported in /transl_except qualifier.
Qualifier	**/trans_splicing**
Definition	Indicates that exons from two RNA molecules are ligated in intermolecular reaction to form mature RNA.
Value format	none
Example	/trans_splicing
Comment	Should be used on features such as CDS, mRNA and other features that are produced as a result of a trans-splicing event. This qualifier should be used only when the splice event is indicated in the "join" operator, eg join(complement(69611.69724),139856..140087).
Qualifier	**/type_material=**
Definition	Indicates that the organism from which this sequence was obtained is a nomenclatural type of the species (or subspecies) corresponding with the /organism identified in the sequence entry.
Value format	"<type-of-type> of <organism name>"
Example	/type_material="type strain of Escherichia coli" /type_material="holotype of Cercopitheus lomamiensis" /type_material="paratype of Cercopitheus lomamiensis"
Comment	<type-of-type> is taken from the INSDC controlled vocabulary for /type_material at: http://www.insdc.org/controlled-vocabulary-typematerial-qualifer <organism name> should be listed as the scientific name (or as a synonym) at the species (or subspecies) node in the taxonomy database. INSDC will automatically populate this qualifier from the NCBI taxonomy database to flag sequences of form type in the INSDC databases (ENA/GenBank/DDBJ).
Qualifier	**/variety=**
Definition	variety (= varietas, a formal Linnaean rank) of organism from which sequence was derived.
Value format	"text"
Example	/variety="insularis"

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.

Comment	Use the cultivar qualifier for cultivated plant varieties, i.e., products of artificial selection; varieties other than plant and fungal variatas should be annotated via /note, e.g. /note="breed:Cukorova".

Campos internos	
guid	Identificador único conformado por el código de la institución, el código del conjunto de datos y el identificador del registro.
modified	Fecha de la última modificación del registro en la base de datos de origen.
_creation	Fecha de creación del registro en la Plataforma de la DGRU.
_creation_user	Nombre de usuario que creó el registro en la Plataforma, en caso de no contar con el dato se utilizará el código del conjunto de datos.
_last_modified	Fecha de última modificación del registro dentro de la Plataforma.
_last_modified_user	Nombre de usuario que realizó la última modificación del registro dentro de la Plataforma de la DGRU.
responsable_quality_control	Nombre de usuario responsable del proceso de Control de Calidad Estructural en la DGRU.

REFERENCIAS

- INSDC (The International Nucleotide Sequence Database Collaboration) (2017). *The DDBJ/ENA/GenBank Feature Table Definition.* Hinxton, Cambridgeshire: EMBL-EBI. Recuperado el 03 de febrero de 2020, de http://www.insdc.org/files/feature_table.html
- NIH (National Institutes of Health) (2019). *GenBank: GenBank Overview.* Bethesda, Maryland: National Center for Biotechnology Information (NCBI). Recuperado de Recuperado el 03 de febrero de 2020, de https://www.ncbi.nlm.nih.gov/genbank/

Sugerencia de cita: DGRU.2020. Estándar para Datos de Secuencias de ADN (GenBank) (Versión 2). Dirección General de Repositorios Universitarios. SDI-UNAM. México.