

The DDBJ/EMBL/GenBank
Feature Table:
Definition

Version 10.0 Dec 2011

DNA Data Bank of Japan, Mishima, Japan.
EMBL-Bank, European Nucleotide Archive, Cambridge, UK.
GenBank, NCBI, Bethesda, MD, USA.

- [1 Introduction](#)
- [2 Overview of the Feature Table format](#)
- [2.1 Format Design](#)
- [2.2 Key aspects of this feature table design](#)
- [2.3 Feature Table Terminology](#)
- [3 Feature table components and format](#)
- [3.1 Naming conventions](#)
- [3.2 Feature keys](#)
- [3.2.1 Purpose](#)
- [3.2.2 Format and conventions](#)
- [3.2.3 Key groups and hierarchy](#)
- [3.2.4 Feature key examples](#)
- [3.3 Qualifiers](#)
- [3.3.1 Purpose](#)
- [3.3.2 Format and conventions](#)
- [3.3.3 Qualifier values](#)
- [3.3.4 Qualifier examples](#)
- [3.4 Location](#)
- [3.4.1 Purpose](#)
- [3.4.2 Format and conventions](#)
- [3.4.3 Location examples](#)
- [4 Feature table Format](#)
- [4.1 Format examples](#)
- [4.2 Definition of line types](#)
- [4.3 Data item positions](#)
- [4.4 Use of blanks](#)
- [5 Examples of sequence annotation](#)
- [5.1 Eukaryotic gene](#)
- [5.2 Bacterial operon](#)
- [5.3 Artificial cloning vector \(circular\)](#)
- [5.4 Plasmid](#)
- [5.5 Repeat element](#)
- [5.6 Immunoglobulin heavy chain](#)
- [5.7 T-cell receptor](#)
- [5.8 Transfer RNA](#)
- [6 Limitations of this feature table design](#)
- [7 Appendices](#)
- [7.1 Appendix I EMBL, GenBank and DDBJ entries](#)
- [7.1.1 EMBL Format](#)
- [7.1.2 GenBank Format](#)
- [7.1.3 DDBJ Format](#)
- [7.2 Appendix II: Feature keys reference](#)
- [7.3 Appendix III: Summary of qualifiers for feature keys](#)
- [7.3.1 Qualifier List](#)
- [7.4 Appendix IV: Controlled vocabularies](#)
- [7.4.1 Nucleotide base codes \(IUPAC\)](#)
- [7.4.2 Modified base abbreviations](#)
- [7.4.3 Amino acid abbreviations](#)
- [7.4.4 Modified and unusual Amino Acids](#)
- [7.4.5 Genetic Code Tables](#)
- [7.4.6 Country Names](#)

1 Introduction

Nucleic acid sequences provide the fundamental starting point for describing and understanding the structure, function, and development of genetically diverse organisms. The GenBank, EMBL, and DDBJ nucleic acid sequence data banks have from their inception used tables of sites and features to describe

the roles and locations of higher order sequence domains and elements within the genome of an organism. In February, 1986, GenBank and EMBL began a collaborative effort (joined by DDBJ in 1987) to devise a common feature table format and common standards for annotation practice.

2 Overview of the Feature Table format

The overall goal of the feature table design is to provide an extensive vocabulary for describing features in a flexible framework for manipulating them. The Feature Table documentation represents the shared rules that allow the three databases to exchange data on a daily basis.

The range of features to be represented is diverse, including regions which:

- * perform a biological function,
- * affect or are the result of the expression of a biological function,
- * interact with other molecules,
- * affect replication of a sequence,
- * affect or are the result of recombination of different sequences,
- * are a recognizable repeated unit,
- * have secondary or tertiary structure,
- * exhibit variation, or have been revised or corrected.

2.1 Format Design

The format design is based on a tabular approach and consists of the following items:

Feature key - a single word or abbreviation indicating functional group

Location - instructions for finding the feature

Qualifiers - auxiliary information about a feature

2.2 Key aspects of this feature table design

* Feature keys allow specific annotation of important sequence features.

* Related features can be easily specified and retrieved.

Feature keys are arranged hierarchically, allowing complex and compound features to be expressed. Both location operators and the feature keys show feature relationships even when the features are not contiguous. The hierarchy of feature keys allows broad categories of biological functionality, such as rRNAs, to be easily retrieved.

* Generic feature keys provide a means for entering new or undefined features. A number of "generic" or miscellaneous feature keys have been added to permit annotation of features that cannot be adequately described by existing feature keys. These generic feature keys will serve as an intermediate step in the identification and addition of new feature keys. The syntax has been designed to allow the addition of new feature keys as they are required.

* More complex locations (fuzzy and alternate ends, for example) can be specified. Each end point of a feature may be specified as a single point, an alternate set of possible end points, a base number beyond which the end point lies, or a region which contains the end point.

* Features can be combined and manipulated in many different ways. The location field can contain operators or functional descriptors specifying what must be done to the sequence to reproduce the feature. For example, a series of exons may be "join"ed into a full coding sequence.

* Standardized qualifiers provide precision and parsibility of descriptive details. A combination of standardized qualifiers and their controlled-vocabulary values enable free-text descriptions to be avoided.

* The nature of supporting evidence for a feature can be explicitly indicated. Features, such as open reading frames or sequences showing sequence similarity to consensus sequences, for which there is no direct experimental evidence can

be annotated. Therefore, the feature table can incorporate contributions from researchers doing computational analysis of the sequence databases. However, all features that are supported by experimental data will be clearly marked as such.

* The table syntax has been designed to be machine parsible. A consistent syntax allows machine extraction and manipulation of sequences coding for all features in the table.

2.3 Feature Table Terminology

The format and wording in the feature table use common biological research terminology whenever possible. For example, an item in the feature table such as:

```
Key          Location/Qualifiers
CDS          23..400
             /product="alcohol dehydrogenase"
             /gene="adhI"
```

might be read as:

The feature CDS is a coding sequence beginning at base 23 and ending at base 400, has a product called 'alcohol dehydrogenase' and is coded for by a gene called "adhI".

A more complex description:

```
Key          Location/Qualifiers
CDS          join(544..589,688..>1032)
             /product="T-cell receptor beta-chain"
```

which might be read as:

This feature, which is a partial coding sequence, is formed by joining elements indicated to form one contiguous sequence encoding a product called T-cell receptor beta-chain.

The following sections contain detailed explanations of the feature table design showing conventions for each component of the feature table, examples of how the format might be implemented, a description of the exact column placement of all the data items and examples of complete sequence entries that have been annotated using the new format. The last section of this document describes known limitations of the current feature table design.

Appendix I gives an example database entry for the DDBJ, GenBank and EMBL formats.

Appendices II and III provide reference manuals for the feature table keys and qualifiers, respectively.

Appendix IV includes controlled vocabularies such as nucleotide base codes, modified base abbreviations, genetic code tables etc.

This document defines the syntax and vocabulary of the feature table. The syntax is sufficiently flexible to allow expression of a single biological entity in numerous ways. In such cases, the annotation staffs at the databases will propose conventions for standard means of denoting the entities. This feature table format is shared by GenBank, EMBL and DDBJ. Comments, corrections, and suggestions may be submitted to any of the database staffs. New format specifications will be added as needed.

3 Feature table components and format

3.1 Naming conventions

Feature table components, including feature keys, qualifiers, accession numbers, database name abbreviations, and location operators, are all named following the same conventions. Component names may be no more than 20 characters long (Feature keys 15, Feature qualifiers 20) and must contain at least one letter. The following characters are permitted to occur in feature table component names:

* Uppercase letters (A-Z)

- * Lowercase letters (a-z) Numbers (0-9)
- * Underscore (_)
- * Hyphen (-)
- * Single quotation mark or apostrophe (')
- * Asterisk (*)

3.2 Feature keys

3.2.1 Purpose

Feature keys indicate

- (1) the biological nature of the annotated feature or
- (2) information about changes to or other versions of the sequence.

The feature key permits a user to quickly find or retrieve similar features or features with related functions.

3.2.2 Format and conventions

There is a defined list of allowable feature keys, which is shown in Appendix II. Each feature must contain a feature key.

3.2.3 Key groups and hierarchy

The feature keys fall into families which are in some sense similar in function and which are annotated in a similar manner. A functional family may have a "generic" or miscellaneous key, which can be recognized by the 'misc.' prefix, that can be used for instances not covered by the other defined keys of that group.

The feature key groups are listed below with a short definition and an annotation example:

1. Difference and change features

Indicate ways in which a sequence should be changed to produce a different

"version":

```
misc_difference location
        /replace="change_location"
```

2. Expression signal features

Indicate regions containing a signal that alters a biological function:

```
misc_signal      location
```

3. Transcript features

Indicate products made by a region:

```
misc_RNA         location
```

4. Binding features

Indicate that a sequence or nucleotide is covalently, non-covalently, or otherwise bound to something else:

```
misc_binding     location
        /bound_moiety="bound molecule"
```

5. Repeat features

Indicate repetitive sequence elements:

```
repeat_region    location
```

6. Recombination features

Indicate regions that have been either inserted or deleted by recombination:

```
misc_recomb      location
```

7. Structure features

Indicate sequence for which there is secondary or tertiary structural information:

misc_structure location

3.2.4 Feature key examples

Key	Description
CDS	Protein-coding sequence
RBS	ribosome binding site
rep_origin	Origin of replication
protein_bind	Protein binding site on DNA
tRNA	mature transfer RNA

See Appendix II for descriptions of all feature keys.

3.3 Qualifiers

3.3.1 Purpose

Qualifiers provide a general mechanism for supplying information about features in addition to that conveyed by the key and location.

3.3.2 Format and conventions

Qualifiers take the form of a slash (/) followed by the qualifier name and, if applicable, an equal sign (=) and a value. Each qualifier should have a single value; if multiple values are necessary, these should be represented by iterating the same qualifier, eg:

Key	Location/Qualifiers
source	1..1000 /culture_collection="ATCC:11775" /culture_collection="CECT:515"

If the location descriptor does not need a continuation line, the first qualifier begins a new line in the feature location column. If the location descriptor requires a continuation line, the first qualifier may follow immediately after the location. Any necessary continuation lines begin in the same column. See Section 4 for a complete description of data item positions.

3.3.3 Qualifier values

Since qualifiers convey many different types of information, there are several value formats:

1. Free text
2. Controlled vocabulary or enumerated values
3. Citation or reference numbers
4. Sequences

3.3.3.1 Free text

Most qualifier values will be a descriptive text phrase which must be enclosed in double quotation marks. When the text occupies more than one line, a single set of quotation marks is required at the beginning and at the end of the text. The text itself may be composed of any printable characters (ASCII values 32-126 decimal). If double quotation marks are used within a free text

string, each set (") must be 'escaped' by placing a second double quotation mark immediately before it ("). For example:

```
/note="This is an example of ""escaped"" quotation marks"
```

3.3.3.2 Controlled vocabulary or enumerated values

Some qualifiers require values from a controlled vocabulary and are entered without quotation marks. For example, the '/direction' qualifier has only three values: 'left', 'right' or 'both'. Qualifier value controlled vocabularies, like feature table component names, must be treated as completely case insensitive: they may be entered and displayed in any combination of upper and lower case ('/direction=Left' '/direction=left' and '/direction=LEFT' are all legal and all convey the same meaning). The database staffs reserve the right to regularize the case of qualifier values. Qualifier value controlled vocabularies will be maintained by the cooperating database staffs. Examples of controlled vocabularies can be found in Appendices IV and V. The database staff should be contacted for the current lists.

3.3.3.3 Citation or reference numbers

The citation or published reference number (as enumerated in the entry 'REFERENCE' or 'RN' data item) should be enclosed in square brackets (e.g., [3]) to distinguish it from other numbers.

3.3.3.4 Sequences

Literal sequence of nucleotide bases e.g., join(12..45,"atgcatt",988..1050) in location descriptors has become illegal starting from implementation of version 2.1 of the Feature Table Definition Document (December 15, 1998)

3.3.4 Qualifier examples

Key	Location/Qualifiers
source	1..1509 /organism="Mus musculus" /strain="CD1" /mol_type="genomic DNA"
promoter	<1..9 /gene="ubc42"
mRNA	join(10..567,789..1320) /gene="ubc42"
CDS	join(54..567,789..1254) /gene="ubc42" /product="ubiquitin conjugating enzyme" /function="cell division control"

3.4 Location

3.4.1 Purpose

The location indicates the region of the presented sequence which corresponds to a feature.

3.4.2 Format and conventions

The location contains at least one sequence location descriptor and may

contain one or more operators with one or more sequence location descriptors. Base numbers refer to the numbering in the entry. This numbering designates the first base (5' end) of the presented sequence as base 1. Base locations beyond the range of the presented sequence may not be used in location descriptors, the only exception being location in a remote entry (see 3.4.2.1, e).

Location operators and descriptors are discussed in more detail below.

3.4.2.1 Location descriptors

The location descriptor can be one of the following:

- (a) a single base number
- (b) a site between two indicated adjoining bases
- (c) a single base chosen from within a specified range of bases (not allowed for new entries)
- (d) the base numbers delimiting a sequence span
- (e) a remote entry identifier followed by a local location descriptor (i.e., a-d)

A site between two adjoining nucleotides, such as endonucleolytic cleavage site, is indicated by listing the two points separated by a caret (^). The permitted formats for this descriptor are n^n+1 (for example 55^56), or, for circular molecules, n^1, where "n" is the full length of the molecule, ie 1000^1 for circular molecule with length 1000.

A single base chosen from a range of bases is indicated by the first base number and the last base number of the range separated by a single period (e.g., '12.21' indicates a single base taken from between the indicated points). From October 2006 the usage of this descriptor is restricted : it is illegal to use "a single base from a range" (c) either on its own or in combination with the "sequence span" (d) descriptor for newly created entries. The existing entries where such descriptors exist are going to be retrofitted.

Sequence spans are indicated by the starting base number and the ending base number separated by two periods (e.g., '34..456'). The '<' and '>' symbols may be used with the starting and ending base numbers to indicate that an end point is beyond the specified base number. The starting and ending base positions can be represented as distinct base numbers ('34..456') or a site between two indicated adjoining bases.

A location in a remote entry (not the entry to which the feature table belongs) can be specified by giving the accession-number and sequence version of the remote entry, followed by a colon ":", followed by a location descriptor which applies to that entry's sequence (i.e. J12345.1:1..15, see also examples below)

3.4.2.2 Operators

The location operator is a prefix that specifies what must be done to the indicated sequence to find or construct the location corresponding to the feature. A list of operators is given below with their definitions and most common format.

complement(location)

Find the complement of the presented sequence in the span specified by "location" (i.e., read the complement of the presented strand in its 5'-to-3' direction)

join(location,location, ... location)

The indicated elements should be joined (placed end-to-end) to form one contiguous sequence

order(location,location, ... location)

The elements can be found in the specified order (5' to 3' direction), but nothing is implied about the reasonableness about joining them

Note : location operator "complement" can be used in combination with either "join" or "order" within the same location; combinations of "join" and "order" within the same location (nested operators) are illegal.

3.4.3 Location examples

The following is a list of common location descriptors with their meanings:

Location	Description
467	Points to a single base in the presented sequence
340..565	Points to a continuous range of bases bounded by and including the starting and ending bases
<345..500	Indicates that the exact lower boundary point of a feature is unknown. The location begins at some base previous to the first base specified (which need not be contained in the presented sequence) and continues to and includes the ending base
<1..888	The feature starts before the first sequenced base and continues to and includes base 888
1..>888	The feature starts at the first sequenced base and continues beyond base 888
102.110	Indicates that the exact location is unknown but that it is one of the bases between bases 102 and 110, inclusive
123^124	Points to a site between bases 123 and 124
join(12..78,134..202)	Regions 12 to 78 and 134 to 202 should be joined to form one contiguous sequence
complement(34..126)	Start at the base complementary to 126 and finish at the base complementary to base 34 (the feature is on the strand complementary to the presented strand)
complement(join(2691..4571,4918..5163))	Joins regions 2691 to 4571 and 4918 to 5163, then complements the joined segments (the feature is on the strand complementary to the presented strand)
join(complement(4918..5163),complement(2691..4571))	Complements regions 4918 to 5163 and 2691 to 4571, then joins the complemented segments (the feature is on the strand complementary to the presented strand)
J00194.1:100..202	Points to bases 100 to 202, inclusive, in the entry (in this database) with primary accession number 'J00194'
join(1..100,J00194.1:100..202)	Joins region 1..100 of the existing entry with the region 100..202 of remote entry J00194

4 Feature table Format

The examples below show the preferred sequence annotations for a number of commonly occurring sequence types. These examples may not be appropriate in all cases but should be used as a guide whenever possible. This section describes the columnar format used to write this feature table in "flat-file" form for distributions of the database.

4.1 Format examples

Feature table format example (EMBL):
 FT source 1..1859


```
FT      /db_xref="taxon:3899"
FT      /organism="Trifolium repens"
FT      /tissue_type="leaves"
FT      /clone_lib="lambda gt10"
FT      /clone="TRE361"
FT      /mol_type="genomic DNA"
FT  CDS  14..1495
FT      /db_xref="MENDEL:11000"
FT      /db_xref="UniProtKB/Swiss-Prot:P26204"
FT      /note="non-cyanogenic"
FT      /EC_number="3.2.1.21"
FT      /product="beta-glucosidase"
FT      /protein_id="CAA40058.1"
FT      /translation="MDFIVAIFALFVISSFTITSTNAVEASTLLDIGNLSR....."
-----+-----+-----+-----+-----+-----+-----+-----+-----+
1         10         20         30         40         50         60         70         79
```

Feature table format example (GenBank):

```
      source      1..8959
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /mol_type="genomic DNA"
      gene        212..8668
                  /gene="NF1"
      CDS         212..8668
                  /gene="NF1"
                  /note="putative"
                  /codon_start=1
                  /product="GAP-related protein"
                  /protein_id="AAA59924.1"
                  /translation="MAAHRPVQAVVSRFDEQLPIKGTGQONHTKVVSTE....."
-----+-----+-----+-----+-----+-----+-----+-----+-----+
1         10         20         30         40         50         60         70         79
```

Feature table format example (DDBJ):

```
      source      1..2136
                  /clone="pK28"
                  /organism="Rattus norvegicus"
                  /strain="Sprague-Dawley"
                  /tissue_type="kidney"
                  /mol_type="genomic DNA"
      mRNA        19..2128
      CDS         31..1212
                  /codon_start=1
                  /function="Dual specificity protein tyrosine/threonine
                  kinase"
                  /product="MAP kinase kinase"
                  /protein_id="BAA02603.1"
                  /translation="MPKKKPTPIQLNPAPDGSVNGTSSAETNLEALQKKL....."
-----+-----+-----+-----+-----+-----+-----+-----+-----+
1         10         20         30         40         50         60         70         79
```

4.2 Definition of line types

The feature table consists of a header line, which contains the column titles for the table, and the individual feature entries. Each feature entry is composed of a feature descriptor line and qualifier and continuation lines, if needed. The feature descriptor line contains the feature's name, key, and location. If the location cannot be contained on the first line of the feature descriptor, it is continued on a continuation line immediately following the descriptor line. If the feature requires further attributes, feature qualifier lines immediately follow the corresponding feature descriptor line (or its continuation). Qualifier information that cannot be contained on one line continues on the following continuation lines as necessary.

Thus, there are 4 types of feature table lines:

Line type	Content	#/entry	#/feature
-----	-----	-----	-----
Header	Column titles	1*	N/A
Feature descriptor	Key and location	1 to many*	1
Feature qualifiers	Qualifiers and values	N/A	0 to many

Continuation lines	Feature descriptor or qualifier continuation	0 to many	0 to many
--------------------	---	-----------	-----------

4.3 Data item positions

The position of the data items within the feature descriptor line is as follows:

column position	data item
-----	-----
1-5	blank
6-20	feature key
21	blank
22-80	location

Data on the qualifier and continuation lines begins in column position 22 (the first 21 columns contain blanks). The EMBL format for all lines differs from the GenBank / DDBJ formats that it includes a line type abbreviation in columns 1 and 2.

4.4 Use of blanks

Blanks (spaces) may, in general, be used within the feature location and qualifier values to make the construction more readable. The following rules should be observed:

- * Names of feature table components may not contain blanks (see Section 3.1)
- * Operator names may not be separated from the following open parenthesis (the beginning of the operand list) by blanks.
- * Qualifiers may not be separated from the preceding slash or the following equals sign (if one) by blanks

5 Examples of sequence annotation

The examples below show the preferred sequence annotations for a number of commonly occurring sequence types. These examples may not be appropriate in all cases but should be used as a guide whenever possible.

5.1 Eukaryotic gene

```

source      1..1509
            /organism="Mus musculus"
            /strain="CD1"
            /mol_type="genomic DNA"
promoter    <1..9
            /gene="ubc42"
mRNA        join(10..567,789..1320)
            /gene="ubc42"
CDS         join(54..567,789..1254)
            /gene="ubc42"
            /product="ubiquitin conjugating enzyme"
            /function="cell division control"
            /translation="MVSSFLLAEYKNLIVNPSEHFKISVNEDNLTEGPPDTLY
            QKIDTVLLSVISLLNEPNPDSPANVDAAKSYRKLYKEDLESYPMEKSLDECS
            AEDIEYFKNVPVNVLPVPSPDDYEDEEMEDGTYILTYDDEDEEDEEMDDE"
exon        10..567
            /gene="ubc42"
            /number=1
intron      568..788
            /gene="ubc42"
            /number=1
exon        789..1320
            /gene="ubc42"
            /number=2

```

polyA_signal 1310..1317
/gene="ubc42"

5.2 Bacterial operon

source 1..9430
/organism="Lactococcus sp."
/strain="MG1234"
/mol_type="genomic DNA"

operon 160..6865
/operon="gal"

-35_signal 160..165
/operon="gal"

-10_signal 179..184
/operon="gal"

CDS 405..1934
/operon="gal"
/gene="galA"
/product="galactose permease"
/function="galactose transporter"

CDS 2003..3001
/operon="gal"
/gene="galM"
/product="aldose 1-epimerase"
/EC_number="5.1.3.3"
/function="mutarotase"

CDS 3235..4537
/operon="gal"
/gene="galK"
/product="galactokinase"
/EC_number="2.7.1.6"

mRNA 189..6865
/operon="gal"

5.3 Artificial cloning vector (circular)

source 1..5300
/organism="Cloning vector pABC"
/lab_host="Escherichia coli"
/mol_type="other DNA"
/focus

source 1..5138
/organism="Escherichia coli"
/mol_type="other DNA"
/strain="K12"

source 5139..5247
/organism="Aequorea victoria"
/mol_type="other DNA"
/dev_stage="adult"

source 5248..5300
/organism="Escherichia coli"
/mol_type="other DNA"
/strain="K12"

CDS join(complement(1..799),complement(5080..5120))
/gene="mob1"
/product="mobilization protein 1"

CDS complement(1697..2512)
/gene="Km"
/product="kanamycin resistance protein"

CDS 3037..3711
/gene="rep1"
/product="replication protein 1"

CDS complement(4170..4829)
/gene="Cm"
/product="chloramphenicol resistance protein"

CDS 5139..5247
/gene="GFP"
/product="green fluorescent protein"

5.4 Plasmid

```

source      1..2245
            /organism="Escherichia coli"
            /plasmid="Plasmid XYZ"
            /strain="K12"
            /mol_type="genomic DNA"
rep_origin  6
            /direction=LEFT
            /note="ori"
CDS         join(complement(567..795),complement(21..349))
            /gene="trbC"
            /product="transfer protein C"
CDS         803..1344
            /gene="traN"
            /product="transfer protein N"
CDS         1559..1985
            /gene="incA"
            /product="incompatability protein A"
CDS         join(2004..2195,3..20)
            /gene="finP"
            /product="fertility inhibition protein P"

```

5.5 Repeat element

```

source      1..1011
            /organism="Homo sapiens"
            /clone="pha281u/1DO"
            /mol_type="genomic DNA"
repeat_region 80..401
            /rpt_type=DISPERSED
            /rpt_family="Alu-J"

```

5.6 Immunoglobulin heavy chain

```

source      1..321
            /organism="Mus musculus "
            /strain="BALB/c2"
            /cell_line="hybridoma 1A4"
            /rearranged
            /mol_type="mRNA"
CDS         <1..>321
            /codon_start=1
            /gene="VFM1-DFL16.1-JH4"
            /product="immunoglobulin heavy chain"
V_region    1..277
            /gene="VFM1"
            /product="immunoglobulin heavy chain variable region"

```

5.7 T-cell receptor

```

source      1..402
            /organism="Homo sapiens"
            /sex="male"
            /cell_type="CD4+ T-lymphocyte"
            /rearranged
            /clone="TCR1A.12"
            /mol_type="mRNA"
sig_peptide 1..54
            /gene="TCR1A"
CDS         1..402
            /gene="TCR1A"

```

```

mat_peptide    /product="T-cell receptor alpha chain"
               55..399
               /gene="TCR1A"
               /product="T-cell receptor alpha chain"
V_region       55..327
               /gene="TCR1A"
J_segment      328..393
               /gene="TCR1A"
C_region       394..399
               /gene="TCR1A"

```

5.8 Transfer RNA

```

source         1..2345
               /organism="Yersinia sp."
               /strain="IP134"
               /mol_type="genomic DNA"
-35_signal     644..650
               /gene="tRNA-Leu(UUR)"
tRNA           655..730
               /gene="tRNA-Leu(UUR)"
               /anticodon=(pos:678..680,aa:Leu)
               /product="transfer RNA-Leu(UUR)"

```

6 Limitations of this feature table design

During the development of the feature table design numerous choices between simplicity and representational power had to be made. In order to create a design which was capable of representing the most common features of biological significance, a certain degree of complexity in the syntax was guaranteed. However, to limit that level of complexity, certain limitations of the design syntax have been accepted.

7 Appendices

7.1 Appendix I EMBL, GenBank and DDBJ entries

7.1.1 EMBL Format

```

ID   X64011; SV 1; linear; genomic DNA; STD; PRO; 756 BP.
XX
AC   X64011; S78972;
XX
SV   X64011.1
XX
DT   28-APR-1992 (Rel. 31, Created)
DT   30-JUN-1993 (Rel. 36, Last updated, Version 6)
XX
DE   Listeria ivanovii sod gene for superoxide dismutase
XX
KW   sod gene; superoxide dismutase.
XX
OS   Listeria ivanovii
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
XX
RN   [1]

```

```

RX  MEDLINE; 92140371.
RA  Haas A., Goebel W.;
RT  "Cloning of a superoxide dismutase gene from Listeria ivanovii by
RT  functional complementation in Escherichia coli and characterization of the
RT  gene product.";
RL  Mol. Gen. Genet. 231:313-322(1992).
XX
RN  [2]
RP  1-756
RA  Kreft J.;
RT  ;
RL  Submitted (21-APR-1992) to the EMBL/GenBank/DDBJ databases.
RL  J. Kreft, Institut f. Mikrobiologie, Universitaet Wuerzburg, Biozentrum Am
RL  Hubland, 8700 Wuerzburg, FRG
XX
FH  Key          Location/Qualifiers
FH
FT  source        1..756
FT                /db_xref="taxon:1638"
FT                /organism="Listeria ivanovii"
FT                /strain="ATCC 19119"
FT                /mol_type="genomic DNA"
FT  RBS           95..100
FT                /gene="sod"
FT  terminator    723..746
FT                /gene="sod"
FT  CDS           109..717
FT                /transl_table=11
FT                /gene="sod"
FT                /EC_number="1.15.1.1"
FT                /db_xref="GOA:P28763"
FT                /db_xref="HSSP:P00448"
FT                /db_xref="InterPro:IPR001189"
FT                /db_xref="UniProtKB/Swiss-Prot:P28763"
FT                /product="superoxide dismutase"
FT                /protein_id="CAA45406.1"
FT                /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTKHHNIYVTKLNEAVSG
FT                HAELASKPGEELVANLDSVP EEEIRGAVRNHGGGHANHTLFWSSSLSPNGGGAPTGNLCAA
FT                IESEFGTFDEFKEKFNA AAAARFGSGWAWLVVNNGKLEIVSTANQDSPLSEGKTPVLGL
FT                DWWEHAYYLKFQNRPEYIDTFWNVINWDERNKRFDAAK"
XX
SQ  Sequence 756 BP; 247 A; 136 C; 151 G; 222 T; 0 other;
    cggtattttaa ggtgttacat agttctatgg aaatagggtc tatacctttc gccttacaat    60
    gtaatttctt .....
//

```

7.1.2 GenBank Format

```

LOCUS      LISOD                      756 bp    DNA        linear    BCT 30-JUN-1993
DEFINITION Listeria ivanovii sod gene for superoxide dismutase.
ACCESSION  X64011 S78972
VERSION    X64011.1 GI:44010
KEYWORDS   sod gene; superoxide dismutase.
SOURCE     Listeria ivanovii
  ORGANISM Listeria ivanovii
            Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE  1 (bases 1 to 756)
  AUTHORS  Haas,A. and Goebel,W.
  TITLE    Cloning of a superoxide dismutase gene from Listeria ivanovii by
            functional complementation in Escherichia coli and characterization
            of the gene product
  JOURNAL  Mol. Gen. Genet. 231 (2), 313-322 (1992)
  MEDLINE  92140371
REFERENCE  2 (bases 1 to 756)
  AUTHORS  Kreft,J.
  TITLE    Direct Submission
  JOURNAL  Submitted (21-APR-1992) J. Kreft, Institut f. Mikrobiologie,
            Universitaet Wuerzburg, Biozentrum Am Hubland, 8700 Wuerzburg, FRG
FEATURES   Location/Qualifiers
  source    1..756
            /organism="Listeria ivanovii"
            /strain="ATCC 19119"
            /db_xref="taxon:1638"
            /mol_type="genomic DNA"
  RBS       95..100

```

```

        /gene="sod"
gene      95..746
        /gene="sod"
CDS       109..717
        /gene="sod"
        /EC_number="1.15.1.1"
        /codon_start=1
        /transl_table=11
        /product="superoxide dismutase"
        /db_xref="GI:44011"
        /db_xref="GOA:P28763"
        /db_xref="InterPro:IPR001189"
        /db_xref="UniProtKB/Swiss-Prot:P28763"
        /protein_id="CAA45406.1"
        /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTKHHNIYVTKLNEAVS
        GHAEELASKPGGEELVANLDSVPPEIRGAVRNHGGGHHANHTLFWSSLSPNGGGAPTGNLK
        AAIESEFGTFDEFKEKFNAAAAARFGSGAWLVVNNGKLEIVSTANQDSPLSEGKTPV
        LGLDVWEHAYYLLKFQNRREYIDTFWNVINWDERNKRFDAAK"
terminator 723..746
        /gene="sod"
ORIGIN
    1 cggtattttaa ggtgttacat agttctatgg aaataggggc tatacctttc gccttacaat
    61 gtaatttctt .....
//

```

7.1.3 DDBJ Format

```

LOCUS      LISOD                      756 bp      DNA      linear      BCT 30-JUN-1993
DEFINITION Listeria ivanovii sod gene for superoxide dismutase.
ACCESSION  X64011 S78972
VERSION    X64011.1  GI:44010
KEYWORDS   sod gene; superoxide dismutase.
SOURCE     Listeria ivanovii
ORGANISM   Listeria ivanovii
            Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE  1 (bases 1 to 756)
AUTHORS    Haas,A. and Goebel,W.
TITLE      Cloning of a superoxide dismutase gene from Listeria ivanovii by
            functional complementation in Escherichia coli and characterization
            of the gene product
JOURNAL    Mol. Gen. Genet. 231 (2), 313-322 (1992)
MEDLINE    92140371
REFERENCE  2 (bases 1 to 756)
AUTHORS    Kreft,J.
TITLE      Direct Submission
JOURNAL    Submitted (21-APR-1992) J. Kreft, Institut f. Mikrobiologie,
            Universitaet Wuerzburg, Biozentrum Am Hubland, 8700 Wuerzburg, FRG
FEATURES   Location/Qualifiers
            source          1..756
                                /organism="Listeria ivanovii"
                                /strain="ATCC 19119"
                                /db_xref="taxon:1638"
                                /mol_type="genomic DNA"
            RBS             95..100
                                /gene="sod"
            gene            95..746
                                /gene="sod"
            CDS              109..717
                                /gene="sod"
                                /EC_number="1.15.1.1"
                                /codon_start=1
                                /transl_table=11
                                /product="superoxide dismutase"
                                /db_xref="GOA:P28763"
                                /db_xref="HSSP:P00448"
                                /db_xref="InterPro:IPR001189"
                                /db_xref="UniProtKB/Swiss-Prot:P28763"
                                /protein_id="CAA45406.1"
                                /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTKHHNIYVTKLNEAVS
                                GHAEELASKPGGEELVANLDSVPPEIRGAVRNHGGGHHANHTLFWSSLSPNGGGAPTGNLK
                                AAIESEFGTFDEFKEKFNAAAAARFGSGAWLVVNNGKLEIVSTANQDSPLSEGKTPV
                                LGLDVWEHAYYLLKFQNRREYIDTFWNVINWDERNKRFDAAK"
            terminator       723..746
                                /gene="sod"

```

```

BASE COUNT      247 a      136 c      151 g      222 t
ORIGIN
    1 cggtattttaa ggtgttacat agttctatgg aaataggggc tatacctttc gccttacaat
   61 gtaatttctt .....
//

```

7.2 Appendix II: Feature keys reference

The following has been organized according to the following format:

Feature Key	the feature key name
Definition	the definition of the key
Mandatory qualifiers	qualifiers required with the key; if there are no mandatory qualifiers, this field is omitted.
Optional qualifiers	optional qualifiers associated with the key
Organism scope	valid organisms for the key; if the scope is any organism, this field is omitted.
Molecule scope	valid molecule types; if the scope is any molecule type, this field is omitted.
References	citations of published reports, usually supporting the feature consensus sequence
Comment	comments and clarifications
Abbreviations:	
accnum	an entry primary accession number
<amino_acid>	abbreviation for amino acid
<base_range>	location descriptor for a simple range of bases
<bool>	Boolean truth value. Valid values are yes and no
<integer>	unsigned integer value
<location>	general feature location descriptor
<modified_base>	abbreviation for modified nucleotide base
[number]	integer representing number of citation in entry's reference list
<repeat_type>	value indicating the organization of a repeated sequence.
"text"	any text or character string. Since the string is delimited by double quotes, double quotes may only appear as part of the string if they appear in pairs. For example, the sentence:

The "label" qualifier is no longer legal.

would be formatted thus:

"The ""label"" qualifier is no longer legal."

Feature Key	assembly_gap
-------------	--------------

Definition	gap between two components of a CON record that is part of a genome 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.)
----------------------	---

Comment	the location span of the assembly_gap feature for an unknown gap is 100 bp, with the 100 bp indicated as 100 "n"'s in sequence.
---------	---

Feature Key	attenuator
-------------	------------

Definition	1) region of DNA at which regulation of termination of transcription occurs, which controls the expression of some bacterial operons;
------------	---

- 2) sequence segment located between the promoter and the first structural gene that causes partial termination of transcription

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) /operon="text" /phenotype="text"
Organism scope	prokaryotes
Molecule scope	DNA
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 /standard_name="text"
Parent Key	CDS
Organism scope	eukaryotes
Feature Key	CAAT_signal
Definition	CAAT box; part of a conserved sequence located about 75 bp up-stream of the start point of eukaryotic transcription units which may be involved in RNA polymerase binding; consensus=GG(C or T)CAATCT [1,2].
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
Molecule scope	DNA
References	[1] Efstratiadis, A. et al. Cell 21, 653-668 (1980) [2] Nevins, J.R. "The pathway of eukaryotic mRNA formation"

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	<pre> /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 /ribosomal_slippage /standard_name="text" /translation="text" /transl_except=(pos:<base_range>,aa:<amino_acid>) /transl_table =<integer> /trans_splicing </pre>
Comment	<pre> /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; </pre>
Feature Key	centromere
Definition	region of biological interest identified as a centromere and which has been experimentally characterized;
Optional qualifiers	<pre> /citation=[number] /db_xref="<database>:<identifier>" /experiment="[CATEGORY:]text" /inference="[CATEGORY:]TYPE[(same species)][:EVIDENCE_BASIS]" /note="text" /standard_name="text" </pre>
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 /standard_name="text"
Parent Key	CDS
Organism scope	eukaryotes
Feature Key	enhancer
Definition	a cis-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter;
Optional qualifiers	/allele="text" /bound_moiety="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"
Organism scope	eukaryotes and eukaryotic viruses
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
 /standard_name="text"

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 same number of
 "n"'s in the sequence.
 No upper or lower limit is set on the size of the gap.

Feature Key GC_signal

Definition GC box; a conserved GC-rich region located upstream of
 the start point of eukaryotic transcription units which
 may occur in multiple copies or in either orientation;
 consensus=GGGCGG;

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 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
/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
/standard_name="text"

```

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 /standard_name="text"
Parent Key	CDS
Organism scope	eukaryotes
Feature Key	LTR
Definition	long terminal repeat, a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses;
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	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 /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 the key RBS is 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 (unsure, 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 /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) /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 /standard_name="text" /trans_splicing
Feature Key	misc_signal
Definition	any region containing a signal controlling or altering gene function or expression that cannot be described by other signal keys (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator, and rep_origin).
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"
/phenotype="text"
/standard_name="text"

```

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 region (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 /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" /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 /standard_name="text" /trans_splicing
Example	/ncRNA_class="miRNA" /ncRNA_class="siRNA" /ncRNA_class="scrRNA"
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 /standard_name="text"
Parent Key	CDS
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 containing genes that encode enzymes that are in the same metabolic pathway and regulatory sequences
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 /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"
/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_signal

Definition recognition region necessary for endonuclease cleavage
 of an RNA transcript that is followed by polyadenylation;
 consensus=AATAAA [1];

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

References [1] Proudfoot, N. and Brownlee, G.G. Nature 263, 211-214
 (1976)

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 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"
/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; includes 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 `promoter`

Definition	region on a DNA molecule involved in RNA polymerase binding to initiate transcription;
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 /standard_name="text"
Molecule scope	DNA
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 RBS is used for ribosome binding sites.
Feature Key	RBS
Definition	ribosome binding site;
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"
References	[1] Shine, J. and Dalgarno, L. Proc Natl Acad Sci USA 71, 1342-1346 (1974) [2] Gold, L. et al. Ann Rev Microb 35, 365-403 (1981)
Comment	in prokaryotes, known as the Shine-Dalgarno sequence: is located 5 to 9 bases upstream of the initiation codon; consensus GGAGGT [1,2].
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" /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 /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 /standard_name="text"
Parent Key	misc_signal
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 /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	/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
/frequency="text"
/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"
/sub_species="text"
/sub_strain="text"
/tissue_lib="text"
/tissue_type="text"
/transgenic
/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 IV /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
Parent key	misc_binding
Comment	STS location to include primer(s) in primer_bind key or primers.
Feature Key	TATA_signal
Definition	TATA box; Goldberg-Hogness box; a conserved AT-rich septamer found about 25 bp before the start point of each eukaryotic RNA polymerase II transcript unit which may be involved in positioning the enzyme for correct initiation; consensus=TATA(A or T)A(A or T) [1,2];
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
Molecule scope	DNA
References	[1] Efstratiadis, A. et al. Cell 21, 653-668 (1980) [2] Corden, J., et al. "Promoter sequences of eukaryotic protein-encoding genes" Science 209, 1406-1414 (1980)
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 region is unique compared to the rest of the chromosome and represent the physical end of the chromosome;

Feature Key terminator

Definition sequence of DNA located either at the end of the transcript that causes RNA polymerase to terminate transcription;

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)
/operon="text"
/standard_name="text"

Molecule scope DNA

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
/standard_name="text"
/tag_peptide=<base_range>

Comment the tmRNA feature key will become valid on 15-Dec-2007

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
/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:<base_range>,aa:<amino_acid>)
 /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
 /standard_name="text"
 /trans_splicing

Feature Key unsure

Definition author is unsure of exact sequence in this region;

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_BASIS]]"
 /locus_tag="text" (single token)

	<pre> /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /standard_name="text" </pre>
Parent Key	CDS
Organism 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	<pre> /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 /standard_name="text" </pre>
Parent Key	CDS
Organism 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	<pre> /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" </pre>
Comment	<p>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;</p> <p>use /replace="" to annotate deletion, e.g.</p> <pre> variation 4..5 /replace="" </pre>
Feature Key	3' UTR

Definition	region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein;
Optional qualifiers	<pre> /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 </pre>
Feature Key	5'UTR
Definition	region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein;
Optional qualifiers	<pre> /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 </pre>
Feature Key	-10_signal
Definition	<p>Pribnow box; a conserved region about 10 bp upstream of the start point of bacterial transcription units which may be involved in binding RNA polymerase;</p> <p>consensus=TAtAaT [1,2,3,4];</p>
Optional qualifiers	<pre> /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) /operon="text" /standard_name="text" </pre>
Organism scope	prokaryotes
Molecule scope	DNA
References	<p>[1] Schaller, H., Gray, C., and Hermann, K. Proc Natl Acad Sci USA 72, 737-741 (1974)</p> <p>[2] Pribnow, D. Proc Natl Acad Sci USA 72, 784-788 (1974)</p> <p>[3] Hawley, D.K. and McClure, W.R. "Compilation and analysis of Escherichia coli promoter DNA sequences" Nucl Acid Res 11, 2237-2255 (1983)</p>

- [4] Rosenberg, M. and Court, D. "Regulatory sequences involved in the promotion and termination of RNA transcription" *Ann Rev Genet* 13, 319-353 (1979)

Feature Key	-35_signal
Definition	a conserved hexamer about 35 bp upstream of the start point of bacterial transcription units; consensus=TTGACa or TGTGACA;
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) /operon="text" /standard_name="text"
Organism scope	prokaryotes
Molecule scope	DNA
References	[1] Takanami, M., et al. <i>Nature</i> 260, 297-302 (1976) [2] Moran, C.P., Jr., et al. <i>Molec Gen Genet</i> 186, 339-346 (1982) [3] Maniatis, T., et al. <i>Cell</i> 5, 109-113 (1975)

7.3 Appendix III: Summary of qualifiers for feature keys

7.3.1 Qualifier List

The following is a list of available qualifiers for feature keys and their usage. The information is arranged as follows:

Qualifier	name of qualifier; qualifier requires a value if followed by an equal sign
Definition	definition of the qualifier
Value format	format of value, if required
Example	example of qualifier with value
Comment	comments, questions and clarifications

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	/anticodon=
Definition	location of the anticodon of tRNA and the amino acid for which it codes
Value format	(pos:<base_range>,aa:<amino_acid>) where base_range is the position of the anticodon and amino_acid is the abbreviation for the amino acid encoded
Example	/anticodon=(pos:34..36,aa:Phe)

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 identifier 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 Multiple /bound_moiety qualifiers are legal on "promoter" and "enhancer" features. 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 which the sequence was obtained
Value format "text"
Example /chromosome="1"

Qualifier /citation=
Definition reference to a citation listed in the entry reference field [integer-number] where integer-number is the number of the reference as enumerated in the reference field
Value format "text"
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 the person who collected the specimen
Value format	"text"
Example	/collected_by="Dan Janzen"
Qualifier	/collection_date=
Definition	date that the specimen was collected
Value format	"DD-Mmm-YYYY", "Mmm-YYYY" or "YYYY"
Example	/collection_date="21-Oct-1952" /collection_date="Oct-1952" /collection_date="1952"
Comment	full date format DD-Mmm-YYYY is preferred; where day and/or month of collection is not known either "Mmm-YYYY" or "YYYY" can be used; three-letter month abbreviation can be one of the following: Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec.
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 localities
Value format	"<country_value>[:<region>][, <locality>]" where country_value is any value from the controlled vocabulary at http://www.insdc.org/country
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 sample 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.
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 habitat.
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". - 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; 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
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"
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: http://www.ncbi.nlm.nih.gov/genome/assembly/agp/AGP_Specification_v2.0.shtml
Qualifier	/gene=
Definition	symbol of the gene corresponding to a sequence region
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 taxonomist who identified the specimen

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:AACN01022672.1', 'InterPro:IPR001900', 'ProDom:PD000600', 'Genscan:2.0', etc. and is structured "[ALGORITHM][:EVIDENCE_DBREF[,EVIDENCE_DBREF]*[,...]]"

Example	<pre> /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" </pre>
Comment	<pre> /inference="non-experimental evidence, no additional details recorded" was used to replace instances of /evidence=NOT_EXPERIMENTAL in December 2005; recommendations for choice of resource acronym for [EVIDENCE_BASIS] are provided in the /inference qualifier vocabulary recommendation document (http://www.insdc.org/inference.html); </pre>
Qualifier	/isolate=
Definition	individual isolate from which the sequence was obtained
Value format	"text"
Example	<pre> /isolate="Patient #152" /isolate="DGGE band PSBAC-13" </pre>
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"
Examples	<pre> /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" </pre>
Comment	<pre> 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. </pre>
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	<pre> /lab_host="Gallus gallus" /lab_host="Gallus gallus embryo" /lab_host="Escherichia coli strain DH5 alpha" /lab_host="Homo sapiens HeLa cells" </pre>
Comment	<pre> 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 </pre>
Qualifier	/lat_lon=
Definition	geographical coordinates of the location where the specimen was collected
Value format	"text"
Example	<pre> /lat_lon="47.94 N 28.12 W" /lat_lon="45.0123 S 4.1234 E" </pre>
Comment	<pre> degrees latitude and longitude in format "d[d.ddd] N S d[dd.ddd] W E" (see the examples) </pre>
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	"paired-ends", "align genus", "align xgenus", "align trnscpt", "within clone", "clone contig", "map", "strobe", "unspecified"
Example	<pre> /linkage_evidence="paired-ends" /linkage_evidence="within clone" </pre>
Comment	<pre> This qualifier is used only for assembly_gap features and its values are controlled by the AGP Specification version 2.0: http://www.ncbi.nlm.nih.gov/genome/assembly/agp/AGP_Specification_v2.0.shtml </pre>

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 /proteid_id identifiers.
Qualifier	/map=
Definition	genomic map position of feature
Value format	"text"
Example	/map="8q12-q13"
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	/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"
Examples	/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.
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 Party Annotation entries; the value "genomic DNA" does not imply that the molecule is nuclear (e.g. organelle and plasmid DNA should be 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
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="scrRNA"
Comment	TYPE is a term taken from the INSDC controlled vocabulary for ncRNA classes (http://www.insdc.org/rna_vocab.html); on 15-Oct-2008, the following terms were valid: <p> "antisense_RNA" "autocatalytically_spliced_intron" "ribozyme" "hammerhead_ribozyme" "RNase_P_RNA" "RNase_MRP_RNA" "telomerase_RNA" "guide_RNA" "rasiRNA" "scrRNA" "siRNA" "miRNA" "piRNA" "snoRNA" "snRNA" "SRP_RNA" "vault_RNA" "Y_RNA" "other" </p> ncRNA classes not yet in the INSDC /ncRNA_class controlled vocabulary can be annotated by entering '/ncRNA_class="other"' with '/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="YP00002"
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"
Comment	currently valid only on Prokaryota-specific features
Qualifier	/organelle=
Definition	type of membrane-bound intracellular structure from which the sequence was obtained
Value format	mitochondrion, nucleomorph, plastid, mitochondrion:kinetoplast, plastid:chloroplast, plastid:apicoplast, plastid:chromoplast, plastid:cyanelle, plastid:leucoplast, plastid:proplastid,
Examples	/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"
Comments	modifier text limited to values from controlled list
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,

	<code>[rev_name: YYY2,]rev_seq: yyyyy2"</code>
Example	<pre> /PCR_primers="fwd_name: C01P1, fwd_seq: ttgatttttttggtcayccwgaagt, rev_name: C01R4, rev_seq: ccwvytardcctarraartgttg" /PCR_primers=" fwd_name: hogel, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i>gtgtatcttact" /PCR_primers="fwd_name: C01P1, fwd_seq: ttgatttttttggtcayccwgaagt, fwd_name: C01P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: C01R4, rev_seq: ccwvytardcctarraartgttg" </pre>
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 Definition	/phenotype= 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 Definition	/pop_variant= name of a variation that characterizes a particular sub-population within a given species. The variation could be in the genotype or the phenotype.
Value format	"text"
Example	<pre> /pop_variant="pop1" /pop_variant="Bear Paw" </pre>
Qualifier Definition	/plasmid= 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 Definition	/product= 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	<pre> /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) </pre>
Qualifier Definition	/protein_id= 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 Definition	/proviral 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 Definition	/pseudo indicates that this feature is a non-functional version of the element named by the feature key

Value format	none
Example	/pseudo
Comment	not to be used for new submissions from 15-APR-2012; After 15-APR-2012 a new qualifier /pseudogene will become valid.
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	/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" /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	organization of repeated sequence
Value format	tandem, inverted, flanking, terminal, direct, dispersed, and other
Example	/rpt_type=INVERTED
Comment	the values are case-insensitive, i.e. both "INVERTED" and "inverted" are valid; Definitions of the values: tandem, a repeat that exists adjacent to another in the same orientation; inverted, a repeat which occurs as part of a set (normally a part) organized in the reverse orientation; flanking, a repeat lying outside the sequence for which it has functional significance (eg. transposon insertion target sites); terminal, a repeat at the ends of and within the sequence for which it has functional significance (eg. transposon LTRs); direct, a repeat that exists not always adjacent but is in the same orientation; dispersed, a repeat that is found dispersed throughout the genome; other, a repeat exhibiting important attributes that cannot be described by other values.
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 tandem 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: Sla" /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.
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".
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"
Examples	/sex="female" /sex="male" /sex="hermaphrodite" /sex="unisexual" /sex="bisexual" /sex="asexual"

Comment	/sex="monoecious" [or monecious]
	/sex="dioecious" [or diecious]
Comment	/sex should be used (instead of /mating_type)
	in the Metazoa, Embryophyta, Rhodophyta & Phaeophyceae;
Comment	/mating_type should be used (instead of /sex)
	in the Bacteria, Archaea & Fungi;
Comment	neither /sex nor /mating_type should be used
	in the viruses;
Comment	outside of the taxa listed above, /mating_type
	should be used unless the value of the qualifier
Comment	is taken from the vocabulary given in the examples
	above
Qualifier	/specimen_voucher=
Definition	identifier for the specimen from which the nucleic acid
Value format	sequenced was obtained
	/specimen_voucher="[<institution-code>:[<collection-code>:]]<specimen_id>"
Example	/specimen_voucher="UAM:Mamm:52179"
	/specimen_voucher="AMCC:101706"
Example	/specimen_voucher="USNM:field series 8798"
	/specimen_voucher="personal:Dan Janzen:99-SRNP-2003"
Example	/specimen_voucher="99-SRNP-2003"
	the /specimen_voucher qualifier is intended to annotate a
Comment	reference to the physical specimen that remains after the
	sequence has been obtained;
Comment	if the specimen was destroyed in the process of sequencing,
	electronic images (e-vouchers) are an adequate substitute for a
Comment	physical voucher specimen; ideally the specimens will be
	deposited in a curated museum, herbarium, or frozen tissue
Comment	collection, but often they will remain in a personal or
	laboratory collection for some time before they are deposited in
Comment	a curated collection;
	there are three forms of specimen_voucher qualifiers; if the
Comment	text of the qualifier includes one or more colons it is a
	'structured voucher'; structured vouchers include
Comment	institution-codes (and optional collection-codes) taken from a
	controlled vocabulary maintained by the INSDC that denotes the
Comment	museum or herbarium collection where the specimen resides;
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
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	/sub_species=
Definition	name of sub-species of organism from which sequence was
Value format	obtained
	"text"
Example	/sub_species="lactis"
Qualifier	/sub_strain=
Definition	name or identifier of a genetically or otherwise modified
Value format	strain from which sequence was obtained, derived from a
	parental strain (which should be annotated in the /strain
Example	qualifier).sub_strain from which sequence was obtained
	"text"
Example	/sub_strain="abis"

Comment	<p>If the parental strain is not given, this should be annotated in the strain qualifier instead of sub_strain. Either:</p> <pre>/strain="K-12" /sub_strain="MG1655" or: /strain="MG1655"</pre>
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	<p>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; the /tag_peptide qualifier (and tmRNA feature) will become valid on 15-Dec-2007</p>
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	<p>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.</p>
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 exceptions in the /transl_except and /codon qualifiers
Value format	IUPAC one-letter amino acid abbreviation, "X" is to be used for AA exceptions.
Example	/translation="MASTFPWPYRGCASTPSLKGLIMCTW"
Comment	<p>to be used with CDS feature only; this is a mandatory qualifier in the CDS feature key except where /pseudo is shown; see /transl_table for definition and location of genetic code Tables.</p>
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	<pre>/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)</pre>
Comment	<p>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)</p>

```

/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;...
Example	/transl_table=4
Comment	genetic code exceptions outside range of specified tables are reported in /codon or /transl_except qualifiers.
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	/variety=
Definition	variety (= varietas, a formal Linnaean rank) of organism from which sequence was derived.
Value format	"text"
Example	/variety="insularis"
Comment	use the cultivar qualifier for cultivated plant varieties, i.e., products of artificial selection; varieties other than plant and fungal varietas should be annotated via /note, e.g. /note="breed:Cukorova"

7.4 Appendix V: Controlled vocabularies

This appendix contains information on the restricted vocabulary fields used in the Feature Table. The information contained in this appendix is subject to change, please contact the database staff for the most recent information concerning controlled vocabularies. This appendix is organized as follows:

Authority	The organization with authority to define the vocabulary
Reference	Publications of (or about) the vocabulary
Contact	Name of database staff responsible for maintaining the database copy of the vocabulary
Scope	Feature Table qualifiers which take members of this vocabulary as values
Listing	A listing of the current vocabulary with definitions or explanations

This appendix includes reference lists for the following controlled vocabulary fields:

- Nucleotide base codes (IUPAC)
- Modified base abbreviations
- Amino acid abbreviations
- Modified and unusual Amino Acids
- Genetic Code Tables
- Country Names

7.4.1 Nucleotide base codes (IUPAC)

Authority	Nomenclature Committee of the International Union of
-----------	--

Biochemistry
 Reference Cornish-Bowden, A. Nucl Acid Res 13, 3021-3030 (1985)
 Contact EMBL-EBI
 Scope Location descriptors

Listing

Symbol	Meaning
-----	-----
a	a; adenine
c	c; cytosine
g	g; guanine
t	t; thymine in DNA; uracil in RNA
m	a or c
r	a or g
w	a or t
s	c or g
y	c or t
k	g or t
v	a or c or g; not t
h	a or c or t; not g
d	a or g or t; not c
b	c or g or t; not a
n	a or c or g or t

7.4.2 Modified base abbreviations

Authority Sprinzl, M. and Gauss, D.H.
 Reference Sprinzl, M. and Gauss, D.H. Nucl Acid Res 10, r1 (1982).
 (note that in Cornish-Bowden, A. Nucl Acid Res 13, 3021-3030 (1985) the IUPAC-IUB declined to recommend a set of abbreviations for modified nucleotides)
 Contact NCBI
 Scope /mod_base

Abbreviation	Modified base description
-----	-----
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
d	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta,D-galactosylqueosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methoxyaminomethyl-2-thiouridine
man q	beta,D-mannosylqueosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-(9-beta-D-ribofuranosyl-2-methyltiopurine-6-yl)carbamoyl)threonine
mt6a	N-(9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
mv	uridine-5-oxyacetic acid-methylester
o5u	uridine-5-oxyacetic acid (v)
osyw	wybutoxosine

p	pseudouridine
q	queosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
t	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u
OTHER	(requires /note= qualifier)

7.4.3 Amino acid abbreviations

Authority	IUPAC-IUB Joint Commission on Biochemical Nomenclature.
Reference	IUPAC-IUB Joint Commission on Biochemical Nomenclature. Nomenclature and Symbolism for Amino Acids and Peptides. Eur. J. Biochem. 138:9-37(1984). IUPAC-IUBMB JCBN Newsletter, 1999 http://www.chem.qmul.ac.uk/iubmb/newsletter/1999/item3.html
Scope	/anticodon, /transl_except
Contact	EMBL-EBI

Listing (note that the abbreviations are legal values for amino acids, not the full names)

Abbreviation	Amino acid name
-----	-----
Ala	A Alanine
Arg	R Arginine
Asn	N Asparagine
Asp	D Aspartic acid (Aspartate)
Cys	C Cysteine
Gln	Q Glutamine
Glu	E Glutamic acid (Glutamate)
Gly	G Glycine
His	H Histidine
Ile	I Isoleucine
Leu	L Leucine
Lys	K Lysine
Met	M Methionine
Phe	F Phenylalanine
Pro	P Proline
Pyl	O Pyrrolysine
Ser	S Serine
Sec	U Selenocysteine
Thr	T Threonine
Trp	W Tryptophan
Tyr	Y Tyrosine
Val	V Valine
Asx	B Aspartic acid or Asparagine
Glx	Z Glutamine or Glutamic acid.
Xaa	X Any amino acid.
Xle	J Leucine or Isoleucine
TERM	termination codon

7.4.4 Modified and unusual Amino Acids

Abbreviation	Amino acid
-----	-----
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid

7.4.5 Genetic Code Tables

Genetic Code [5]

Invertebrate Mitochondrial Code (transl_table=5)

```
AAs      = FFLSSSSSY**CWWWLLLLPPPHHQRRRIIMTTTTNNKKSSSVVVVAAAADDEEGGGG
Starts   = ---M-----MMMM-----M---
Base1    = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCCCAAAAAAGGGGGGGGGGGGGG
Base2    = TTTTCCCAAAGGGGTTCCTCCAAGGGGTTCCTCCAAGGGGTTCCTCCAAGGGG
Base3    = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

Genetic Code [6]

Ciliate, Dasycladacean and Hexamita Nuclear Code (transl_table=6)

```
AAs      = FFLSSSSSYQCC*WLLLLPPPHHQRRRIIIMTTTTNNKKSSRRVVVAAAAADDEEGGGG
Starts   = -----M-----
Base1    = TTTTNTTTTTTTTTTTTTCCCCCCCCCCCCCCCCCCAAAAAAAAAAAGGGGGGGGGGGGGGGGGGG
Base2    = TTTTCCCCAAAGGGGTTCCTCCCAAAGGGGTTCCTCCCAAAGGGGTTCCTCCCAAAGGGGG
Base3    = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

Genetic Code [9]

Echinoderm and Flatworm Mitochondrial Code (transl_table=9)

```
AAs      = FFLSSSSSY**CWWLLLLPPPHHQRRRIIIMTTTTNNNKSSSVVVVAAAADDEEGGGG
Starts   = -----M-----M-----
Base1    = TTTTNTTTTTTTTTTTCCCCCCCCCCCCCAAAAAAAAAAAGGGGGGGGGGGGGG
Base2    = TTTTCCCAAAGGGGTTTCCCAAAGGGGTTTCCCAAAGGGGTTTCCCAAAGGGG
Base3    = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

Genetic Code [10]

Euplotid Nuclear Code (transl_table=10)

```
AAs      = FFLSSSSSY**CCCWLLLLPPPHHQRRRIIIMTTTTNNKKSSRRVVVVAADDEEGGGG
Starts   = -----M-----
Base1    = TTTTNTTTTTTTTTTTTTCCCCCCCCCCCCCAAAAAAAAAAAGGGGGGGGGGGGGGGGGGG
Base2    = TTTTCCCAAAGGGGTTTCCCAAAGGGGTTTCCCAAAGGGGTTTCCCAAAGGGG
Base3    = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

Genetic Code [11]

Bacterial, Archaeal and Plant Plastid Code (transl_table=11)

[illegible]

Genetic Code [12]

Alternative Yeast Nuclear Code (transl table=12)

```
AAs      = FFLSSSSSY**CC*WLLSPPPPHHQRRRIIIMTTTTNNKKSSRRVVVVAADDEEGGGG
Starts   = -----M-----M-----
Base1    = TTTTNTTTTTTTTTTTTTCCCCCCCCCCCCCAAAAAAAAAAAAAAGGGGGGGGGGGGGGGGG
Base2    = TTTTCCCCAAAGGGGTTCCTCCCAAAGGGGTTCCTCCCAAAGGGGTTCCTCCCAAAGGGG
Base3    = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

Genetic Code [13]

Ascidian Mitochondrial Code (transl table=13)

```
AAs      = FFLSSSSSY**CWWLLLLPPPHHQRRRIIMTTTTNNKKSSGGVVVVAADDEEGGGG
Starts   = ---M-----MM-----M-----
Base1    = TTTT-----TTTTCCCCCCCCCCCCCCAAAAAAAAAAGGGGGGGGGGGGG
Base2    = TTTTCCCAAAAGGGGTTTCCCAAAAGGGGTTCCTCCCAAAAGGGGTTTCCCAAAAGGGG
Base3    = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

Genetic Code [14]

Alternative Flatworm Mitochondrial Code (transl table=14)

```
AAs      = FFLSSSSYY*CWWLLLLPPPHHQRRRIIIMTTTTNNKSSSVVVVAAAADDEEGGGG
Starts   = -----M-----
Base1    = TTTTNTTTTTTTTTTTTCCCCCCCCCCCCCAAAAAAAAAAGGGGGGGGGGGGGGGGGGG
Base2    = TTTTCCCAAAGGGGTTCCTCCCAAAGGGGTTCCTCCCAAAGGGGTTCCTCCCAAAGGGG
Base3    = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

Blepharisma Nuclear Code (transl_table=15)

Chlorophycean Mitochondrial Code (transl_table=16)

Trematode Mitochondrial Code (transl_table=21)

Scenedesmus obliquus mitochondrial

Thraustochytrium Mitochondrial Code (transl_table=23)

Pterobranchia mitochondrial code (transl_table=24)

60/60