

Features allowed in INSDC sequence files.

These are the features that are allowed in DDBJ/EMBL/GenBank format files. Any other features are not valid in those files.

Feature	Description
assembly_gap	gap between two components of a genome or transcriptome assembly

Feature	Description
---------	-------------

C_region	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
----------	---

Feature	Description
CDS	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.

Feature
Key Description

centromere region
of
bi-
o-
log-
i-
cal
in-
ter-
est
iden-
ti-
fied
as
a
cen-
tromere
and
which
has
been
ex-
per-
i-
men-
tally
characterized

Feature	
Key	Description

D-loop	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
--------	---

Feature	Description
D_segment	Diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
exon	region of genome that codes for portion of spliced mRNA, rRNA and tRNA; may contain 5'UTR, all CDSs and 3' UTR

Feature	
Key	Description
gap	gap in the sequence
gene	region of bi- o- log- i- cal in- ter- est iden- ti- fied as a gene and for which a name has been assigned

Feature	Description
iDNA	intervening DNA; DNA which is eliminated through any of several kinds of recombination

Feature	Description
---------	-------------

intron	segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
--------	--

Feature	Description
J_segment	segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains

Feature
Key Description

mat_peptide
peptide
or
protein
coding
sequence;
coding
sequence
for
the
mature
or
final
peptide
or
protein
product
following
post-translational
modifi-
cation;
the
location
does
not
include
11 the
stop
codon
(un-
like
the
cor-
re-

Feature	Description
misc_bind	binding 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)

Feature	
Key	Description
misc_difference	sequence is different from that presented in the entry and cannot be described by any other difference key (old_sequence, variation, or modified_base)

Feature	Description
misc	<p>feature</p> <p>of</p> <p>bi-</p> <p>o-</p> <p>log-</p> <p>i-</p> <p>cal</p> <p>in-</p> <p>ter-</p> <p>est</p> <p>which</p> <p>can-</p> <p>not</p> <p>be</p> <p>de-</p> <p>scribed</p> <p>by</p> <p>any</p> <p>other</p> <p>fea-</p> <p>ture</p> <p>key;</p> <p>a</p> <p>new</p> <p>or</p> <p>rare</p> <p>feature</p>

Feature
Key Description

misc siteomb

of
any
gen-
er-
al-
ized,
site-
specific
or
replica-
tive
re-
com-
bi-
na-
tion
event
where
there
is
a
break-
age
and
re-
union
of
du-
plex
DNA
that
can-
not
be
de-
scribed
by
other
re-
com-
bi-
na-
tion
keys

15or

qual-
i-
fiers
of
source
key
(/proviral)

Feature	Description
misc_RNA	RNA 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)

Feature	Description
misc_structure	sec- ondary or ter- tiary nu- cleotide struc- ture or con- for- ma- tion that can- not be de- scribed by other Struc- ture keys (stem_loop and D- loop)
mobile_element	of genome con- tain- ing mo- bile elements

Feature	Description
modified_base	indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)

Feature	Description
mRNA	messenger RNA; includes 5'untranslated region (5'UTR), coding sequences (CDS, exon) and 3'untranslated region (3'UTR)

Feature	Description
---------	-------------

ncRNA	
-------	--

	non-protein-coding gene, other than ribosomal RNA and transfer RNA, the functional molecule of which is the RNA transcript
--	--

N_region	intranucleotides inserted between rearranged immunoglobulin segments.
----------	---

Feature	Description
old_sequence	pre- sented se- quence re- vises a pre- vi- ous ver- sion of the se- quence at this location

Feature	Description
operon	region containing polycistronic transcript including a cluster of genes that are under the control of the same regulatory sequences/promoter and in the same biological pathway

Feature	Description
oriT	origin of transfer; region of a DNA molecule where transfer is initiated during the process of conjugation or mobilization

Feature	Description
polyA site	on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation

Feature	Description
precursor RNA	RNA species that is not yet the mature RNA product; may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR)

Feature	Description
prim_transcript	(initial, unprocessed) transcript; may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR)

Feature	
Key	Description

primerobind	covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic e.g., PCR primer elements
-------------	---

Feature	Description
propeptide	peptide cod- ing se- quence; cod- ing se- quence for the do- main of a pro- pro- tein that is cleaved to form the ma- ture pro- tein product.
protein	bind covalent pro- tein bind- ing site on nu- cleic acid

Feature
Key Description

regulatory
re-
gion
of
se-
quence
that
func-
tions
in
the
reg-
u-
la-
tion
of
tran-
scrip-
tion,
trans-
la-
tion,
repli-
ca-
tion
or
chro-
matin
structure

repeating
region
of
genome
con-
tain-
ing
re-
peat-
ing
units

Feature	Description
---------	-------------

rep_origin	of replication; starting site for duplication of nucleic acid to give two identical copies
------------	--

Feature	Description
---------	-------------

rRNAmature	ri- bo- so- mal RNA; RNA com- po- nent of the ri- bonu- cle- o- pro- tein par- ti- cle (ri- bo- some) which as- sem- bles amino acids into proteins.
------------	--

Feature	Description
S_region	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

Feature	Description
sig_peptide	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

Feature
Key Description

sourceidentifies
the
bi-
o-
log-
i-
cal
source
of
the
spec-
i-
fied
span
of
the
se-
quence;
this
key
is
manda-
tory;
more
than
one
source
key
per
se-
quence
is
al-
lowed;
ev-
ery
en-
try/record
will
have,
as
a
min-
i-
mum,
34ei-
ther
a
sin-
gle
source
key
span-

Feature	Description
---------	-------------

stem loop;	
------------	--

a
double-
helical
re-
gion
formed
by
base-
pairing
be-
tween
ad-
ja-
cent
(in-
verted)
com-
ple-
men-
tary
se-
quences
in
a
sin-
gle
strand
of
RNA
or
DNA.

Feature	
Key	Description
STS	<p>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</p>

Feature	Description
telomere	region of bi-ological interest identified as a telomere and which has been experimentally characterized

Feature
Key Description

tmRNA transfer

mes-
sen-
ger
RNA;
tm-
RNA
acts
as
a
tRNA
first,
and
then
as
an
mRNA
that
en-
codes
a
pep-
tide
tag;
the
ri-
bo-
some
trans-
lates
this
mRNA
re-
gion
of
tm-
RNA
and
at-
taches
the
en-
coded
pep-
tide
38tag
to
the
C-
terminus
of
the
un-

Feature	Description
---------	-------------

transit peptide	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
-----------------	---

Feature	Description
---------	-------------

tRNAmature	trans-fer RNA, a small RNA molecule (75-85 bases long) that mediates the trans-lation of a nucleic acid sequence into an amino acid sequence
------------	--

Feature
Key Description

unsure
small
re-
gion
of
se-
quenced
bases,
gen-
er-
ally
10
or
fewer
in
its
length,
which
could
not
be
con-
fi-
dently
iden-
ti-
fied.
Such
a
re-
gion
might
con-
tain
called
bases
(A,
T,
G,
or
C),
or
a
mix-
ture
41 of
called-
bases
and
uncalled-
bases
(‘N’).
The

Feature	Description
V_region	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

Feature	Description
V_segment	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

Feature	Description
---------	-------------

variation	re- lated strain con- tains sta- ble mu- ta- tions from the same gene (e.g., RFLPs, poly- mor- phisms, etc.) which dif- fer from the pre- sented se- quence at this lo- ca- tion (and pos- si- bly others)
-----------	--

Feature
Key Description

3'UTR)

re-
gion
at
the
3'
end
of
a
ma-
ture
tran-
script
(fol-
low-
ing
the
stop
codon)
that
is
not
trans-
lated
into
a
pro-
tein;
2)
re-
gion
at
the
3'
end
of
an
RNA
virus
(fol-
low-
ing
the
last
stop
45codon)
that
is
not
trans-
lated
into
a

Feature	Description
---------	-------------

5'UTR)	re- gion at the 5' end of a ma- ture tran- script (pre- ced- ing the ini- tia- tion codon) that is not trans- lated into a pro- tein; 2) re- gion at the 5' end of an RNA virus genome (pre- ced- ing 46the first ini- tia- tion codon) that is
--------	--

Feature
Key Description
