

## 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 Key	Description
assembly_gap	gap between two com- po- nents of a genome or tran- scrip- tome assembly

<hr/>	
Feature	
Key	Description
<hr/>	
C_region	constant re- gion of im- munoglob- u- lin light and heavy chains, and T- cell re- cep- tor al- pha, beta, and gamma chains; in- cludes one or more ex- ons de- pend- ing on the par- tic- u- lar chain

Feature	
Key	Description
CDS	coding se- quence; se- quence of nu- cleotides that cor- re- sponds with the se- quence of amino acids in a pro- tein (lo- ca- tion in- cludes stop codon); fea- ture in- cludes amino acid con- cep- tual translation.

<hr/>	
Feature	
Key	Description
<hr/>	
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 re- gion within mi- to- chon- drial DNA in which a short stretch of RNA is paired with one strand of DNA, dis- plac- ing the orig- i- nal part- ner DNA strand in this re- gion; also used to de- scribe 5 the dis- place- ment of a re- gion
--------	--

Feature	
Key	Description
D_diversity	Diversity
D_seg	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

<hr/>	
Feature	
Key	Description
<hr/>	
iDNA	intervening DNA; DNA which is elim- i- nated through any of sev- eral kinds of recombination



<hr/>	
Feature	
Key	Description
<hr/>	
introna	<p>segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it</p>

<hr/>	
Feature	
Key	Description
<hr/>	
J_segjoining	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 the stop codon (un- like the cor- re-
-------------	---

<hr/>	
Feature	
Key	Description
<hr/>	
misc_binding	binding in nu- cleic acid which co- va- lently or non- covalently binds an- other moi- ety that can- not be de- scribed by any other bind- ing key (primer_bind or protein_bind)

<hr/>	
Feature	
Key	Description
<hr/>	
misc_difference	<p>sequence is dif- fer- ent from that pre- sented in the en- try and can- not be de- scribed by any other dif- fer- ence key (old_sequence, vari- a- tion, or modified_base)</p>

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

Feature	
Key	Description

misc	<div> <div>site</div> <div>comb</div> <div>of</div> <div>any</div> <div>gen-</div> <div>er-</div> <div>al-</div> <div>ized,</div> <div>site-</div> <div>specific</div> <div>or</div> <div>replica-</div> <div>tive</div> <div>re-</div> <div>com-</div> <div>bi-</div> <div>na-</div> <div>tion</div> <div>event</div> <div>where</div> <div>there</div> <div>is</div> <div>a</div> <div>break-</div> <div>age</div> <div>and</div> <div>re-</div> <div>union</div> <div>of</div> <div>du-</div> <div>plex</div> <div>DNA</div> <div>that</div> <div>can-</div> <div>not</div> <div>be</div> <div>de-</div> <div>scribed</div> <div>by</div> <div>other</div> <div>re-</div> <div>com-</div> <div>bi-</div> <div>na-</div> <div>tion</div> <div>keys</div> </div>
15	<div> <div>or</div> <div>qual-</div> <div>i-</div> <div>fiers</div> <div>of</div> <div>source</div> <div>key</div> <div>(/proviral)</div> </div>

Feature	
Key	Description
misc_RNA	RNA transcript or RNA product that cannot be defined by other RNA keys (prim_transcript, pre-cur- sor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, tran- sit_peptide, mat_peptide, in- tron, polyA_site, ncRNA, rRNA and tRNA)



<hr/>	
Feature	
Key	Description
<hr/>	
misc_structure	secondary or tertiary nucleotide structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
mobile_element	region of genome containing mobile elements

<hr/>	
Feature	
Key	Description
<hr/>	
modified_base	the base indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)

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

Feature	
Key	Description
ncRNA	non-protein-coding gene, other than ribosomal RNA and transfer RNA, the functional molecule of which is the RNA transcript
N_region	intronic nucleotides inserted between rearranged immunoglobulin segments.

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

Feature	
Key	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

<hr/>	
Feature	
Key	Description
<hr/>	
oriT	origin of trans- fer; re- gion of a DNA molecule where trans- fer is ini- ti- ated dur- ing the pro- cess of con- ju- ga- tion or mobilization

<hr/>	
Feature	
Key	Description
<hr/>	
polyA site	<p>site on an RNA tran- script to which will be added ade- nine residues by post- transcriptional polyadenylation</p>



Feature	
Key	Description
precursor	any 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)

<hr/>	
Feature	
Key	Description
<hr/>	
prim_transcript	<p>transcript (initial, un-processed) transcript; may include ncRNA, rRNA, tRNA, 5' un-translated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' un-translated region (3'UTR)</p>

<hr/>	
Feature	
Key	Description
<hr/>	
primer	bind
	covalent
	primer
	bind-
	ing
	site
	for
	ini-
	tia-
	tion
	of
	repli-
	ca-
	tion,
	tran-
	scrip-
	tion,
	or
	re-
	verse
	tran-
	scrip-
	tion;
	in-
	cludes
	site(s)
	for
	syn-
	thetic
	e.g.,
	PCR
	primer
	elements

<hr/>	
Feature	
Key	Description
<hr/>	
propeptide	propeptide 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	protein covalent pro- tein bind- ing site on nu- cleic acid

<hr/>	
Feature	
Key	Description
<hr/>	
regulatory	any 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

<hr/>	
Feature	
Key	Description
<hr/>	
rep_origin	origin of repli- ca- tion; start- ing site for du- pli- ca- tion of nu- cleic acid to give two iden- ti- cal copies

<hr/>	
Feature	
Key	Description
<hr/>	
rRNA	Amature
	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	
Key	Description
S_region	switch re- gion of im- munoglob- u- lin heavy chains; in- volved in the re- ar- range- ment of heavy chain DNA lead- ing to the ex- pres- sion of a dif- fer- ent im- munoglob- u- lin class from the same B- cell



<hr/>	
Feature	
Key	Description
<hr/>	
sig_peptide	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

Feature	
Key	Description

source identifies

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,

34 ei-

ther  
a  
sin-  
gle  
source  
key  
span-

<hr/>	
Feature	
Key	Description
<hr/>	
stem_loopin;	<p>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.</p>

Feature	
Key	Description
STS	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

<hr/>	
Feature	
Key	Description
<hr/>	
telomeric region	of bi- o- log- i- cal in- ter- est iden- ti- fied as a telom- ere and which has been ex- per- i- men- tally characterized

Feature	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	
Key	Description
transit peptide	peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organelar protein; this domain is involved in post-translational import of the protein into the organelle

<hr/>	
Feature	
Key	Description
<hr/>	
tRNA	transfer RNA, a small RNA molecule (75-85 bases long) that mediates the translation 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	
Key	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	
Key	Description
V_segmentable	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	
Key	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	
Key	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  
the  
first  
ini-  
tia-  
tion  
codon)  
that  
is

<hr/>	
Feature	
Key	Description
<hr/>	