

ST.26 - ANNEX I

CONTROLLED VOCABULARY

Final Draft

TABLE OF CONTENTS

SECTION 1: LIST OF NUCLEOTIDES.....	2
SECTION 2: LIST OF MODIFIED NUCLEOTIDES.....	2
SECTION 3: LIST OF AMINO ACIDS.....	4
SECTION 4: LIST OF MODIFIED AND UNUSUAL AMINO ACIDS.....	5
SECTION 5: FEATURE KEYS FOR NUCLEIC SEQUENCES.....	6
SECTION 6: QUALIFIERS FOR NUCLEIC SEQUENCES.....	21
SECTION 7: FEATURE KEYS FOR AMINO ACID SEQUENCES.....	40
SECTION 8: QUALIFIERS FOR AMINO ACID SEQUENCES.....	47
SECTION 9: GENETIC CODE TABLES.....	48

SECTION 1: LIST OF NUCLEOTIDES

The nucleotide base codes to be used in sequence listings are presented in Table 1. The symbol “t” will be construed as thymine in DNA and uracil in RNA when it is used with no further description. Where an ambiguity symbol (representing two or more bases in the alternative) is appropriate, the most restrictive symbol should be used. For example, if a base in a given position could be “a or g,” then “r” should be used, rather than “n”. The symbol “n” will be construed as “a or c or g or t/u” when it is used with no further description.

Table 1: List of nucleotides

Symbol	Nucleotide
a	adenine
c	cytosine
g	guanine
t	thymine in DNA/uracil in RNA (t/u)
m	a or c
r	a or g
w	a or t/u
s	c or g
y	c or t/u
k	g or t/u
v	a or c or g; not t/u
h	a or c or t/u; not g
d	a or g or t/u; not c
b	c or g or t/u; not a
n	a or c or g or t/u; “unknown” or “other”

SECTION 2: LIST OF MODIFIED NUCLEOTIDES

The abbreviations listed in Table 2 are the only permitted values for the mod_base qualifier. Where a specific modified nucleotide is not present in the table below, then the abbreviation “OTHER” must be used as its value. If the abbreviation is “OTHER,” then the complete unabbreviated name of the modified base must be provided in a note qualifier. The abbreviations provided in Table 2 must not be used in the sequence itself.

Table 2: List of modified nucleotides

Abbreviation	Modified Nucleotide
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
dhu	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta-D-galactosylqueuosine
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
m4c	N4-methylcytosine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine

Abbreviation	Modified Nucleotide
mam5u	5-methylaminomethyluridine
mam5s2u	5-methylaminomethyl-2-thiouridine
man q	beta-D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
mv	uridine-5-oxoacetic acid-methylester
o5u	uridine-5-oxyacetic acid (v)
osyw	wybutoxosine
p	pseudouridine
q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
m5u	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)

SECTION 3: LIST OF AMINO ACIDS

The amino acid codes to be used in sequence are presented in Table 3. Where an ambiguity symbol (representing two or more amino acids in the alternative) is appropriate, the most restrictive symbol should be used. For example, if an amino acid in a given position could be aspartic acid or asparagine, the symbol "B" should be used, rather than "X". The symbol "X" will be construed as any one of "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", or "V", when it is used with no further description.

Table 3: List of amino acids

Symbol	Amino acid
A	Alanine
R	Arginine
N	Asparagine
D	Aspartic acid (Aspartate)
C	Cysteine
Q	Glutamine
E	Glutamic acid (Glutamate)
G	Glycine
H	Histidine
I	Isoleucine
L	Leucine
K	Lysine
M	Methionine
F	Phenylalanine
P	Proline
O	Pyrrolysine
S	Serine
U	Selenocysteine
T	Threonine
W	Tryptophan
Y	Tyrosine
V	Valine
B	Aspartic acid or Asparagine
Z	Glutamine or Glutamic acid
J	Leucine or Isoleucine
X	A or R or N or D or C or Q or E or G or H or I or L or K or M or F or P or O or S or U or T or W or Y or V; "unknown" or "other"

SECTION 4: LIST OF MODIFIED AND UNUSUAL AMINO ACIDS

Table 4 lists the only permitted abbreviations for a modified or unusual amino acid in the mandatory qualifier “NOTE” for feature keys “MOD_RES” or “SITE”. The value for the qualifier “NOTE” must be either an abbreviation from this table, where appropriate, or the complete, unabbreviated name of the modified amino acid. The abbreviations (or full names) provided in this table must not be used in the sequence itself.

Table 4: List of modified and unusual amino acids

Abbreviation	Modified or Unusual 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
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
alle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
Melle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

SECTION 5: FEATURE KEYS FOR NUCLEIC SEQUENCES

This paragraph contains the list of allowed feature keys to be used for nucleotide sequences, and lists mandatory and optional qualifiers. The feature keys are listed in alphabetic order. The feature keys can be used for either DNA or RNA unless otherwise indicated under "Molecule scope". Certain Feature Keys may be appropriate for use with artificial sequences in addition to the specified "organism scope".

Feature key names must be used in the XML instance of the sequence listing exactly as they appear following "Feature key" in the descriptions below, except for the feature keys 3'UTR and 5'UTR. See "Comment" in the description for the 3'UTR and 5'UTR feature keys.

5.1.	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 gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes
5.2.	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 may include amino acid conceptual translation
	Optional qualifiers	allele codon_start EC_number exception function gene gene_synonym map note number operon product protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_table trans_splicing
	Comment	codon_start qualifier 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 Standard or universal genetic code table; genetic code exceptions outside the range of the specified tables are reported in transl_except qualifier; only one of the qualifiers translation and pseudo are permitted with a CDS feature key; when the translation qualifier is used, the protein_id qualifier is mandatory if the translation product contains four or more specifically defined amino acids

5.3.	Feature Key	centromere
	Defini ti on	region of bi ologi cal i nterest i denti fi ed as a centromere and which has been experimental ly characteri zed
	Opti onal qual i fi ers	note standard_name
	Comment	the centromere feature describes the interval of DNA that corresponds to a region where chromatids are held and a kinetochore is formed

5.4.	Feature Key	D-loop
	Defini ti on	di splacement loop; a region wi thi n mi tochondri al DNA i n whi ch a short stretch of RNA is pai red wi th one strand of DNA, di splaci ng the ori gi nal partne r DNA strand i n thi s regi on; al so used to descri be the di splacement of a regi on of one strand of duplex DNA by a si ngle stranded i nvader i n the reacti on catal yzed by RecA protei n
	Opti onal qual i fi ers	al l e l e gene gene_synonym map note
	Mol ecul e scope	DNA

5.5.	Feature Key	D_segment
	Defini ti on	Di versity segment of i mmunoglobul i n heavy chain, and T-cell receptor beta chain
	Opti onal qual i fi ers	al l e l e gene gene_synonym map note product pseudo pseudogene standard_name
	Organi sm scope	eukaryotes

5.6.	Feature Key	exon
	Defini ti on	regi on of genome that codes for porti on of spliced mRNA, rRNA and tRNA; may contai n 5' UTR, al l CDSs and 3' UTR
	Opti onal qual i fi ers	al l e l e EC_number functi on gene gene_synonym map note number product pseudo pseudogene standard_name trans_spl i ci ng

5.7.	Feature Key	gene
	Definition	region of biological interest identified as a gene and for which a name has been assigned
	Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene phenotype standard_name 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 its positions at the ends; it is meant to represent a region where the gene is located.

5.8.	Feature Key	iDNA
	Definition	intervening DNA; DNA which is eliminated through any of several kinds of recombination
	Optional qualifiers	allele function gene gene_synonym map note number standard_name
	Molecule scope	DNA
	Comment	e.g., in the somatic processing of immunoglobulin genes.

5.9.	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 function gene gene_synonym map note number pseudo pseudogene standard_name trans_splicing

5.10.	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 gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes
5.11.	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 EC_number function gene gene_synonym map note product pseudo pseudogene standard_name
5.12.	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
	Optional qualifiers	allele function gene gene_synonym map note
	Comment	note that the regulatory feature key and regulatory_class qualifier with the value "ribosome_binding_site" should be used for ribosome binding sites

5.13.	Feature Key	mi_sc_difference
	Definition	featured sequence differs from the presented sequence at this location and cannot be described by any other Difference key (variation, or modified_base)
	Optional qualifiers	allele clone compare gene gene_synonym map note phenotype replace standard_name
	Comment	the mi_sc_difference feature key should be used to describe variability introduced artificially, e.g. by genetic manipulation or by chemical synthesis; use the replace qualifier to annotate a deletion, insertion, or substitution.

5.14.	Feature Key	mi_sc_feature
	Definition	region of biological interest which cannot be described by any other feature key; a new or rare feature
	Optional qualifiers	allele function gene gene_synonym map note number phenotype product pseudo pseudogene standard_name
	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

5.15.	Feature Key	mi_sc_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 (provi ral)
	Optional qualifiers	allele gene gene_synonym map note standard_name
	Molecule scope	DNA

5.16.	Feature Key	mi sc_RNA
	Defini tion	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, transi t_peptide, mat_peptide, intron, polyA_site, ncRNA, rRNA and tRNA)
	Opti onal qual i fiers	al l e l e functi on gene gene_synonym map note operon product pseudo pseudogene standard_name trans_spl i ci ng
5.17.	Feature Key	mi sc_structure
	Defini tion	any secondary or tertiary nucleotide structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
	Opti onal qual i fiers	al l e l e functi on gene gene_synonym map note standard_name
5.18.	Feature Key	mo bil e_el ement
	Defini tion	region of genome containing mo bil e elements
	Mandatory qual i fiers	mo bil e_el ement_type
	Opti onal qual i fiers	al l e l e functi on gene gene_synonym map note rpt_fami l y rpt_type standard_name
5.19.	Feature Key	modi fi ed_base
	Defini tion	the indicated nucleotide is a modi fi ed nucleotide and shoul d be substi tuted for by the indicated molecule (given in the mod_base qual i fier value)
	Mandatory qual i fiers	mod_base
	Opti onal qual i fiers	al l e l e frequency gene gene_synonym map note
	Comment	value for the mandatory mod_base qual i fier is limited to the restricted vocabulary for modi fi ed base abbreviations in Section 2 of this Annex.

5.20.	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 function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.21.	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
	Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
	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
5.22.	Feature Key	N_region
	Definition	extra nucleotides inserted between rearranged immunoglobulin segments
	Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes

5.23.	Feature Key	operon
	Defini tion	region contain ing polyci stronic transcript including a cluster of genes that are under the control of the same regulatory sequences/promotor and in the same biological pathway
	Mandatory qual i fiers	operon
	Opti onal qual i fiers	al l e l e functi on map note phenotype pseudo pseudogene standard_name

5.24.	Feature Key	ori T
	Defini tion	origin of transfer; region of a DNA molecule where transfer is initiated during the process of conjugation or mobilization
	Opti onal qual i fiers	al l e l e bound_moi ety di recti on gene gene_synonym map note rpt_fami l y rpt_type rpt_uni t_range rpt_uni t_seq standard_name
	Mol ecul e Scope	DNA
	Comment	rep_origi n should be used for origins of replication; direction qualifier has legal values RIGHT, LEFT and BOTH, however only RIGHT and LEFT are valid when used in conjunction with the ori T feature; origins of transfer can be present in the chromosome; plasmids can contain multiple origins of transfer

5.25.	Feature Key	pol yA_si te
	Defini tion	si te on an RNA transcript to which wi l l be added adenine residues by post-transcripti onal pol yadenyl ati on
	Opti onal qual i fiers	al l e l e gene gene_synonym map note
	Organi sm scope	eukaryotes and eukaryoti c vi ruses

5.26.	Feature Key	precursor_RNA
	Definition	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)
	Optional qualifiers	allele function gene gene_synonym map note operon product standard_name 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
5.27.	Feature Key	prim_transcript
	Definition	primary (initial, unprocessed) transcript; may include ncRNA, rRNA, tRNA, 5' untranslated region (5' UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3' UTR)
	Optional qualifiers	allele function gene gene_synonym map note operon standard_name
5.28.	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 gene gene_synonym map note standard_name
	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; since PCR reactions most often involve pairs of primers, a single primer_bind key may use the order(location, location) operator with two locations, or a pair of primer_bind keys may be used

5.29.	Feature Key	protein_bind
	Definition	non-covalent protein binding site on nucleic acid
	Mandatory qualifiers	bound_moiety
	Optional qualifiers	allele function gene gene_synonym map note operon standard_name
	Comment	note that the regulatory feature key and regulatory_class qualifier with the value "ribosome_binding_site" should be used for ribosome binding sites
5.30.	Feature Key	regulatory
	Definition	any region of a sequence that functions in the regulation of transcription or translation;
	Mandatory qualifiers	regulatory_class
	Optional qualifiers	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name
5.31.	Feature Key	repeat_region
	Definition	region of genome containing repeating units
	Optional qualifiers	allele function gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq satellite
5.32.	standard_nameFeature Key	rep_origin
	Definition	origin of replication; starting site for duplication of nucleic acid to give two identical copies
	Optional Qualifiers	allele direction gene gene_synonym map note standard_name
	Comment	direction qualifier has valid values: RIGHT, LEFT, or BOTH

5.33.	Feature Key	rRNA
	Definition	mature ribosomal RNA; RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
	Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name
	Comment	rRNA sizes should be annotated with the product qualifier

5.34.	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 gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes

5.35.	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 function gene gene_synonym map note product pseudo pseudogene standard_name

5.36.	Feature Key	source
	Definition	identifies the source of the sequence; this key is mandatory; every sequence will have a single source key spanning the entire sequence
	Mandatory qualifiers	organism mol_type
	Optional qualifiers	cell_line cell_type chromosome clone clone_lib collected_by collection_date cultivar dev_stage ecotype environmental_sample germline haplogroup haplotype host identified_by isolate isolation_source lab_host lat_lon macronuclear map mating_type note organelle PCR_primers plasmid pop_variant provi ral rearranged segment serotype serovar sex strain sub_clone sub_species sub_strain tissue_lib tissue_type variety
	Molecule scope	any

5.37.	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 function gene gene_synonym map note operon standard_name

5.38.	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 gene gene_synonym map note standard_name
	Molecule scope	DNA
	Comment	STS location to include primer(s) in primer_bind key or primers
5.39.	Feature Key	telomere
	Definition	region of biological interest identified as a telomere and which has been experimentally characterized
	Optional qualifiers	note rpt_type rpt_unit_range rpt_unit_seq standard_name
	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 represents the physical end of the chromosome
5.40.	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 function gene gene_synonym map note product pseudo pseudogene standard_name tag_peptide
5.41.	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 function gene gene_synonym map note product pseudo pseudogene standard_name

5.42.	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 anticodon function gene gene_synonym map note product pseudo pseudogene standard_name trans_splicing
5.43.	Feature Key	unsure
	Definition	a small region of sequenced bases, generally 10 or fewer in its length, which could not be confidently identified. Such a region might contain called bases (a, t, g, or c), or a mixture of called-bases and uncalled-bases ('n').
	Optional qualifiers	allele compare gene gene_synonym map note replace
	Comment	use the replace qualifier to annotate a deletion, insertion, or substitution.
5.44.	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 gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes
5.45.	Feature Key	V_segment
	Definition	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
	Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes

5.46.	Feature Key	variation
	Definition	a related strain contains stable mutations from the same gene (e.g., RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
	Optional qualifiers	allele compare frequency gene gene_synonym map note phenotype product replace standard_name
	Comment	used to describe alleles, RFLP's, and other naturally occurring mutations and polymorphisms; variability arising as a result of genetic manipulation (e.g. site directed mutagenesis) should be described with the misc_difference feature; use the replace qualifier to annotate a deletion, insertion, or substitution
5.47.	Feature Key	3' UTR
	Definition	1) region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein; 2) region at the 3' end of an RNA virus (following the last stop codon) that is not translated into a protein;
	Optional qualifiers	allele function gene gene_synonym map note standard_name trans_splicing
	Comment	The apostrophe character has special meaning in XML, and must be substituted with "'" in the value of an element. Thus "3' UTR" must be represented as "3'UTR" in the XML file, i.e., <INSDFeature_key>3'UTR</INSDFeature_key>.
5.48.	Feature Key	5' UTR
	Definition	1) region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein; 2) region at the 5' end of an RNA virus (preceding the first initiation codon) that is not translated into a protein;
	Optional qualifiers	allele function gene gene_synonym map note standard_name trans_splicing
	Comment	The apostrophe character has special meaning in XML, and must be substituted with "'" in the value of an element. Thus "5' UTR" must be represented as "5'UTR" in the XML file, i.e., <INSDFeature_key>5'UTR</INSDFeature_key>.

SECTION 6: QUALIFIERS FOR NUCLEIC SEQUENCES

This section contains the list of qualifiers to be used for features in nucleotide sequences. The qualifiers are listed in alphabetic order.

Where a Value format of "none" is indicated in the description of a qualifier (e.g. germline), the INSDQualifier_value element must not be used.

6.1.	Qualifier	allele
	Definition	name of the allele for the given gene
	Value format	free text
	Example	<INSDQualifier_value>adh1-1</INSDQualifier_value>
	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.
6.2.	Qualifier	anticodon
	Definition	Location of the anticodon of tRNA and the amino acid for which it codes
	Value format	(pos: <location>, aa: <amino_acid>, seq: <text>) where location is the position of the anticodon and <amino_acid> is the three letter abbreviation for the amino acid encoded and seq is the sequence of the anticodon
	Example	<INSDQualifier_value>(pos: 34..36, aa: Phe, seq: aaa)</INSDQualifier_value> <INSDQualifier_value>(pos: join(5, 495..496), aa: Leu, seq: taa)</INSDQualifier_value> <INSDQualifier_value>(pos: complement(4156..4158), aa: Glu, seq: ttg)</INSDQualifier_value>
6.3.	Qualifier	bound_moiety
	Definition	name of the molecule/complex that may bind to the given feature
	Value format	free text
	Example	<INSDQualifier_value>GAL4</INSDQualifier_value>
	Comment	A single bound_moiety qualifier is legal on the "misc_binding", "oriT" and "protein_bind" features.
6.4.	Qualifier	cell_line
	Definition	cell line from which the sequence was obtained
	Value format	free text
	Example	<INSDQualifier_value>MCF7</INSDQualifier_value>
6.5.	Qualifier	cell_type
	Definition	cell type from which the sequence was obtained
	Value format	free text
	Example	<INSDQualifier_value>leukocyte</INSDQualifier_value>

6.6.	Qual i fi er	chromosome
	Defi ni ti on	chromosome (e.g. Chromosome number) from which the sequence was obtained
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>1</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>X</INSDQual i fi er_val ue>
6.7.	Qual i fi er	cl one
	Defi ni ti on	cl one from which the sequence was obtained
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>l ambda-hl L7. 3</INSDQual i fi er_val ue>
	Comment	not more than one cl one should be speci fi ed for a given source feature; where the sequence was obtained from mul ti ple clones it may be further described in the feature table using the feature key misc_feature and a note qual i fi er to speci fy the mul ti ple clones.
6.8.	Qual i fi er	cl one_lib
	Defi ni ti on	cl one library from which the sequence was obtained
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>l ambda-hl L7</INSDQual i fi er_val ue>
6.9.	Qual i fi er	codon_start
	Defi ni ti on	i ndicates the offset at which the fi rst complete codon of a coding feature can be found, relative to the fi rst base of that feature.
	Val ue format	1 or 2 or 3
	Exampl e	<INSDQual i fi er_val ue>2</INSDQual i fi er_val ue>
6.10.	Qual i fi er	col l ected_by
	Defi ni ti on	name of persons or institute who collected the specimen
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>Dan Janzen</INSDQual i fi er_val ue>
6.11.	Qual i fi er	col l ecti on_date
	Defi ni ti on	date that the specimen was collected.
	Val ue format	DD-Mmm-YYYY, Mmm-YYYY or YYYY
	Exampl e	<INSDQual i fi er_val ue>21-Oct-1952</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>Oct-1952</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>1952</INSDQual i fi er_val ue>
	Comment	full date format DD-Mmm-YYYY is preferred; where day and/or month of collection is not known ei ther "Mmm-YYYY" or "YYYY" can be used; three-letter month abbreviation can be one of the fol l owing: Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec.

6.12.	Qualifier	compare
	Definition	Reference details of an existing public INSD entry to which a comparison is made
	Value format	[accession-number.sequence-version]
	Example	<INSDQualifier_val ue>AJ634337.1</INSDQualifier_val ue>
	Comment	This qualifier may be used on the following features: misc_difference, unsure, and variation. 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.
6.13.	Qualifier	cultivar
	Definition	cultivar (cultivated variety) of plant from which sequence was obtained
	Value format	free text
	Example	<INSDQualifier_val ue>Nipponbare</INSDQualifier_val ue> <INSDQualifier_val ue>Tenuifolius</INSDQualifier_val ue> <INSDQualifier_val ue>Candy Cane</INSDQualifier_val ue> <INSDQualifier_val ue>IR36</INSDQualifier_val ue>
	Comment	'cultivar' is applied solely to products of artificial selection; use the variety qualifier for natural, named plant and fungal varieties.
6.14.	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	free text
	Example	<INSDQualifier_val ue>fourth instar larva</INSDQualifier_val ue>
6.15.	Qualifier	direction
	Definition	direction of DNA replication
	Value format	left, right, or both where left indicates toward the 5' end of the sequence (as presented) and right indicates toward the 3' end
	Example	<INSDQualifier_val ue>left</INSDQualifier_val ue>
	Comment	The values left, right, and both are permitted when the direction qualifier is used to annotate a rep_origin feature key. However, only left and right values are permitted when the direction qualifier is used to annotate an oriT feature key.

6.16.	Qualifier	EC_number
	Definition	Enzyme Commission number for enzyme product of sequence
	Value format	free text
	Example	<INSDQualifier_val ue>1.1.2.4</INSDQualifier_val ue> <INSDQualifier_val ue>1.1.2.-</INSDQualifier_val ue> <INSDQualifier_val ue>1.1.2.n</INSDQualifier_val ue>
	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.
6.17.	Qualifier	ecotype
	Definition	a population within a given species displaying genetically based, phenotypic traits that reflect adaptation to a local habitat
	Value Format	free text
	Example	<INSDQualifier_val ue>Col umbi a</INSDQualifier_val ue>
	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.
6.18.	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
	Comment	used only with the source feature key; source feature keys containing the environmental_sample qualifier should also contain the isolation_source qualifier. Sequences including environmental_sample must not include the strain qualifier.
6.19.	Qualifier	exception
	Definition	indicates that the coding region cannot be translated using standard biological rules
	Value format	One of the following controlled vocabulary phrases: RNA editing rearrangement required for product annotated by transcript or proteomic data
	Example	<INSDQualifier_val ue>RNA edi ting</INSDQualifier_val ue> <INSDQualifier_val ue>rearrangement requi red for product</INSDQualifier_val ue>
	Comment	only to be used to describe biological mechanisms such as RNA editing; protein translation of a CDS with an exception qualifier will be different from the according conceptual translation; must not be used where transl_except qualifier would be adequate, e.g. in case of stop codon completion use.

6.20.	Qualifier	frequency
	Definition	frequency of the occurrence of a feature
	Value format	free text representing the proportion of a population carrying the feature expressed as a fraction
	Example	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.21.	Qualifier	function
	Definition	function attributed to a sequence
	Value format	free text
	Example	<INSDQualifier_value>essential for recognition of cofactor </INSDQualifier_value>
	Comment	The function qualifier is used when the gene name and/or product name do not convey the function attributable to a sequence.
6.22.	Qualifier	gene
	Definition	symbol of the gene corresponding to a sequence region
	Value format	free text
	Example	<INSDQualifier_value>iIvE</INSDQualifier_value>
	Comment	Use gene qualifier to provide the gene symbol; use standard_name qualifier to provide the full gene name.
6.23.	Qualifier	gene_synonym
	Definition	synonymous, replaced, obsolete or former gene symbol
	Value format	free text
	Example	<INSDQualifier_value>Hox-3.3</INSDQualifier_value> in a feature where the gene qualifier value is Hoxc6
	Comment	used where it is helpful to indicate a gene symbol synonym; when used, a primary gene symbol must always be indicated in a gene qualifier
6.24.	Qualifier	germline
	Definition	the sequence presented 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
	Comment	germline qualifier should not be used to indicate that the source of the sequence is a gamete or germ cell; germline and rearranged qualifiers cannot be used in the same source feature; germline and rearranged qualifiers 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 qualifiers should not be used outside of the Craniata (taxid=89593)

6.25.	Qual i fi er	hapl ogroup
	Defi ni ti on	name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups.
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>H*</INSDQual i fi er_val ue>
6.26.	Qual i fi er	hapl otype
	Defi ni ti on	name for a specific set 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.
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>Dw3 B5 Cw1 A1</INSDQual i fi er_val ue>
6.27.	Qual i fi er	host
	Defi ni ti on	natural (as opposed to laboratory) host to the organism from which sequenced molecule was obtained
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>Homo sapi ens</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>Homo sapi ens 12 year old girl</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>Rhi zobi um NGR234</INSDQual i fi er_val ue>
6.28.	Qual i fi er	i denti fi ed_by
	Defi ni ti on	name of the expert who identified the specimen taxonomically
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>John Burns</INSDQual i fi er_val ue>
6.29.	Qual i fi er	i sol ate
	Defi ni ti on	individual isolate from which the sequence was obtained
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>Pati ent #152</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>DGGE band PSBAC-13</INSDQual i fi er_val ue>
6.30.	Qual i fi er	i sol ati on_source
	Defi ni ti on	describes the physical, environmental and/or local geographical source of the biological sample from which the sequence was derived
	Val ue format	free text
	Exampl es	<INSDQual i fi er_val ue>rumen isolates from standard Pelleted ration-fed steer #67</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>permanent Antarctic sea ice</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>denitrifying activated sludge from carbon_limited continuous reactor</INSDQual i fi er_val ue>
	Comment	used only with the source feature key; source feature keys containing an environmental_sample qual i fi er should also contain an i sol ati on_source qual i fi er

6.31.	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	free text
	Example	<INSDQualifier_value>Galus gallus</INSDQualifier_value> <INSDQualifier_value>Galus gallus embryo</INSDQualifier_value> <INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value> <INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value>
	Comment	the full binomial scientific name of the host organism should be used when known; extra conditional information relating to the host may also be included
6.32.	Qualifier	lat_lon
	Definition	geographical coordinates of the location where the specimen was collected
	Value format	free text - degrees latitude and longitude in format "d[.ddd] N S d[.ddd] W E"
	Example	<INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value> <INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>
6.33.	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
6.34.	Qualifier	map
	Definition	genomic map position of feature
	Value format	free text
	Example	<INSDQualifier_value>8q12-q13</INSDQualifier_value>
6.35.	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	free text
	Examples	<INSDQualifier_value>MAT-1</INSDQualifier_value> <INSDQualifier_value>plus</INSDQualifier_value> <INSDQualifier_value>-</INSDQualifier_value> <INSDQualifier_value>odd</INSDQualifier_value> <INSDQualifier_value>even</INSDQualifier_value>
	Comment	mating_type qualifier values male and female are valid in the prokaryotes, but not in the eukaryotes; for more information, see the entry for the sex qualifier.

6.36.	Qual i fier	mobile_element_type
	Defi ni ti on	type and name or identifier of the mobile element which is described by the parent feature
	Val ue 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
	Exampl e	<INSBQual i fier_val ue>transposon:Tnp9</INSBQual i fier_val ue>
	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" for <mobile_element_type> requires a <mobile_element_name>
6.37.	Qual i fier	mod_base
	Defi ni ti on	abbreviation for a modified nucleotide base
	Val ue format	modified base abbreviation chosen from this Annex, Section 2
	Exampl e	<INSBQual i fier_val ue>m5c</INSBQual i fier_val ue> <INSBQual i fier_val ue>OTHER</INSBQual i fier_val ue>
	Comment	specific modified nucleotides not found in Section 2 of this Annex are annotated by entering OTHER as the value for the mod_base qualifier and including a note qualifier with the full name of the modified base as its value
6.38.	Qual i fier	mol_type
	Defi ni ti on	molecule type of sequence
	Val ue format	One chosen from the following: genomic DNA genomic RNA mRNA tRNA rRNA other RNA other DNA transcribed RNA viral cRNA unassigned DNA unassigned RNA
	Exampl e	<INSBQual i fier_val ue>genomic DNA</INSBQual i fier_val ue> <INSBQual i fier_val ue>other RNA</INSBQual i fier_val ue>
	Comment	mol_type qualifier is mandatory on the source feature key; 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; 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.

6.39.	Qualifier	ncRNA_class
	Definition	a structured description of the classification of the non-coding RNA described by the ncRNA parent key
	Value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: anti_sense_RNA autocatalytically_spliced_intron ribozyme hammerhead_ribozyme lncRNA RNase_P_RNA RNase_MRP_RNA telomerase_RNA guide_RNA rasiRNA scrRNA siRNA miRNA piRNA snoRNA snRNA SRP_RNA vault_RNA Y_RNA other
	Example	<INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>siRNA</INSDQualifier_value> <INSDQualifier_value>scrRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value>
	Comment	specific ncRNA types not yet in the ncRNA_class controlled vocabulary can be annotated by entering "other" as the ncRNA_class qualifier value, and providing a brief explanation of novel ncRNA_class in a note qualifier
6.40.	Qualifier	note
	Definition	any comment or additional information
	Value format	free text
	Example	<INSDQualifier_value>A comment about the feature</INSDQualifier_value>
6.41.	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	free text (with no whitespace characters)
	Example	<INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value>
	Comment	text limited to integers, letters or combination of integers and/or letters represented as a data value that contains no whitespace characters; any additional terms should be included in a standard_name qualifier. Example: a number qualifier with a value of 2A and a standard_name qualifier with a value of "long"

6.42.	Qual i fi er	operon
	Defi ni ti on	name of the group of contiguous genes transcribed into a single transcript to which that feature belongs
	Val ue format	free text
	Exampl e	<INSDDQual i fi er_val ue>lac</INSDDQual i fi er_val ue>
6.43.	Qual i fi er	organelle
	Defi ni ti on	type of membrane-bound intracellular structure from which the sequence was obtained
	Val ue format	One of the following controlled vocabulary terms and phrases: chromatophore hydrogenosome mitochondrion nucl eomorph plastid mitochondrion: kinetoplast plastid: chloroplast plastid: apicoplast plastid: chromoplast plastid: cyanoblast plastid: leucoplast plastid: proplastid
	Exampl es	<INSDDQual i fi er_val ue>chromatophore</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>hydrogenosome</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>mitochondrion</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>nucl eomorph</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>plastid</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>mitochondrion: kinetoplast</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>plastid: chloroplast</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>plastid: apicoplast</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>plastid: chromoplast</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>plastid: cyanoblast</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>plastid: leucoplast</INSDDQual i fi er_val ue> <INSDDQual i fi er_val ue>plastid: proplastid</INSDDQual i fi er_val ue>
6.44.	Qual i fi er	organism
	Defi ni ti on	scientific name of the organism that provided the sequenced genetic material, if known, or the available taxonomic information if the organism is unclassified; or an indication that the sequence is a synthetic construct
	Val ue format	free text
	Exampl e	<INSDDQual i fi er_val ue>Homo sapiens</INSDDQual i fi er_val ue>

6.45.	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	[fwd_name: XXX1,]fwd_seq: xxxxx1, [fwd_name: XXX2,]fwd_seq: xxxxx2, [rev_name: YYY1,]rev_seq: yyyyy1, [rev_name: YYY2,]rev_seq: yyyyy2
	Example	<INSDQualifier_value>fwd_name: C01P1, fwd_seq: ttgatttttggtcayccwgaagt, rev_name: C01R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value> <INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i> ; i > ; gtgtatcttact</INSDQualifier_value> <INSDQualifier_value>fwd_name: C01P1, fwd_seq: ttgatttttggtcayccwgaagt, fwd_name: C01P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: C01R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value>
	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 symbols from Section 1 of this Annex, except for the modified bases; those must be enclosed within angle brackets < >. In XML, the angle brackets < and > must be substituted with < and > since they are reserved characters in XML.
6.46.	Qualifier	phenotype
	Definition	phenotype conferred by the feature, where phenotype is defined as a physical, biochemical or behavioural characteristic or set of characteristics
	Value format	free text
	Example	<INSDQualifier_value>erythromycin resistance</INSDQualifier_value>
6.47.	Qualifier	plasmid
	Definition	name of naturally occurring plasmid from which the sequence was obtained, where plasmid is defined as an independently replicating genetic unit that cannot be described by chromosome or segment qualifiers
	Value format	free text
	Example	<INSDQualifier_value>pC589</INSDQualifier_value>
6.48.	Qualifier	pop_variant
	Definition	name of subpopulation or phenotype of the sample from which the sequence was derived
	Value format	free text
	Example	<INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>
6.49.	Qualifier	product
	Definition	name of the product associated with the feature, e.g. the mRNA of an mRNA feature, the polypeptide of a CDS, the mature peptide of a mat_peptide, etc.
	Value format	free text
	Example	<INSDQualifier_value>trypsinogen</INSDQualifier_value> (when qualifier appears in CDS feature) <INSDQualifier_value>trypsin</INSDQualifier_value> (when qualifier appears in mat_peptide feature) <INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (when qualifier appears in mRNA feature)

6. 50.	Qual i fi er	protei n_id
	Defi ni ti on	protei n sequence identi fication number, an integer used in a sequence listing to designate the protei n sequence encoded by the codi ng sequence identi fied in the correspondi ng CDS feature key and translati on qual i fi er
	Val ue format	an integer greater than zero
	Exampl e	<INSDQual i fi er_val ue>89</INSDQual i fi er_val ue>
6. 51.	Qual i fi er	provi ral
	Defi ni ti on	this qual i fi er is used to flag sequence obtai ned from a vi rus or phage that is i ntegrated i nto the genome of another organi sm
	Val ue format	none
6. 52.	Qual i fi er	pseudo
	Defi ni ti on	i ndicates that this feature is a non-functi onal versi on of the element named by the feature key
	Val ue format	none
	Comment	The qual i fi er pseudo should be used to describe non-functi onal genes that are not formall y descri bed as pseudogenes, e.g. CDS has no translati on due to other reasons than pseudogeni sation events. Other reasons may i nclude sequenci ng or assembly errors. In order to annotat e pseudogenes the qual i fi er pseudogene must be used, i ndicati ng the TYPE of pseudogene.
6. 53.	Qual i fi er	pseudogene
	Defi ni ti on	i ndicates that this feature is a pseudogene of the element named by the feature key
	Val ue format	TYPE where TYPE is one of the followi ng control led vocabul ary terms or phrases: processed unprocessed uni tary allelic unknown
	Exampl e	<INSDQual i fi er_val ue>processed</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>unprocessed</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>uni tary</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>al lelic</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>unknown</INSDQual i fi er_val ue>
	Comment	Defi ni ti ons of TYPE values: processed - the pseudogene has arisen by reverse transcrip ti on of a mRNA i nto cDNA, followed by rei ntegrati on i nto the genome. Therefore, it has lost any i ntron/exon structure, and it might have a pseudo-polyA-tail. unprocessed - the pseudogene has arisen from a copy of the parent gene by duplicati on followed by accumulati on of random mutat i ons. The changes, compared to their functi onal homolog, i nclude i nserti ons, deleti ons, premature stop codons, frameshi fts and a higher proporti on of non-synonym ous versus synonym ous substi tuti ons. uni tary - the pseudogene has no parent. It is the original gene, which is functi onal i n some speci es but disrupted i n some way (i ndels, mutat i on, recomb i nati on) i n another speci es or strain. allelic - a (uni tary) pseudogene that is stable i n the populati on but i mportantl y it has a functi onal alternati ve allele also i n the populati on. i .e., one strain may have the gene, another strain may have the pseudogene. MHC haplotypes have allelic pseudogenes. unknown - the submi tter does not know the method of pseudogeni zati on.

6. 54.	Qual i f i e r	rearranged
	Def i n i t i o n	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
	Val u e f o r m a t	none
	Co m m e n t	The rearranged qual i f i e r should not be used to annotate chromosome rearrangements that are not involved in an adaptive immune response; germline and rearranged qual i f i e r s cannot be used in the same source feature; germline and rearranged qual i f i e r s 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 qual i f i e r s should not be used outside of the Craniata (taxid=89593)
6. 55.	Qual i f i e r	regulatory_class
	Def i n i t i o n	a structured description of the classification of transcriptional and translational regulatory elements in a sequence
	Val u e f o r m a t	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: attenuator CAAT_si gnal enhancer enhancer_b l o c k i n g _ e l e m e n t GC_si gnal i m p r i n t i n g _ c o n t r o l _ r e g i o n i n s u l a t o r l o c u s _ c o n t r o l _ r e g i o n m i n u s _ 3 5 _ s i g n a l m i n u s _ 1 0 _ s i g n a l response_e l e m e n t p o l y A _ s i g n a l _ s e q u e n c e promoter r i b o s o m e _ b i n d i n g _ s i t e r i b o s o m e _ b i n d i n g _ s i t e s i l e n c e r TATA_box terminator other
	Exa m p l e	<INSBQual i f i e r _ v a l u e>promoter</INSBQual i f i e r _ v a l u e> <INSBQual i f i e r _ v a l u e>enhancer</INSBQual i f i e r _ v a l u e> <INSBQual i f i e r _ v a l u e>r i b o s o m e _ b i n d i n g _ s i t e</INSBQual i f i e r _ v a l u e>
	Co m m e n t	spec i f i c regulatory classes not yet in the regulatory_class controlled vocabulary can be annotated by entering "other" as the regulatory_class qual i f i e r value and providing a brief explanation of the novel regulatory_class in a note qual i f i e r
6. 56.	Qual i f i e r	replace
	Def i n i t i o n	i n d i c a t e s t h a t t h e s e q u e n c e i d e n t i f i e d i n a f e a t u r e ' s l o c a t i o n i s r e p l a c e d b y t h e s e q u e n c e s h o w n i n t h e q u a l i f i e r ' s v a l u e ; i f n o s e q u e n c e (i . e . , n o v a l u e) i s c o n t a i n e d w i t h i n t h e q u a l i f i e r , t h i s i n d i c a t e s a d e l e t i o n
	Val u e f o r m a t	free text
	Exa m p l e	<INSBQual i f i e r _ v a l u e>a</INSBQual i f i e r _ v a l u e> <INSBQual i f i e r _ v a l u e></INSBQual i f i e r _ v a l u e> - for a del et i o n

6.57.	Qual i fi er	ri bosomal_sl i ppage
	Defi ni ti on	during protein translation, certain sequences can program ribosomes to change to an alternative reading frame by a mechanism known as ribosomal slippage
	Val ue format	none
	Com ment	a join operator, e.g.: [join(486..1784,1787..4810)] should be used in the CDS spans to indicate the location of ribosomal_slippage
6.58.	Qual i fi er	rpt_fami ly
	Defi ni ti on	type of repeated sequence; "Alu" or "Kpn", for example
	Val ue format	free text
	Exampl e	<INSQQual i fi er_val ue>Al u</INSQQual i fi er_val ue>
6.59.	Qual i fi er	rpt_type
	Defi ni ti on	structure and distribution of repeated sequence
	Val ue format	One of the following controlled vocabulary terms or phrases: tandem direct inverted flanking nested dispersed terminal long_terminal_repeat non_ltr_retrotransposon_polymeric_tract centromeric_repeat telomeric_repeat x_element_combinatorial_repeat y_prime_element other
	Exampl e	<INSQQual i fi er_val ue>inverted</INSQQual i fi er_val ue> <INSQQual i fi er_val ue>long_terminal_repeat</INSQQual i fi er_val ue>
	Com ment	Definitions of the values: tandem - a repeat that exists adjacent to another in the same orientation; direct - a repeat that exists not always adjacent but is in the same orientation; inverted - a repeat pair occurring in reverse orientation to one another on the same molecule; flanking - a repeat lying outside the sequence for which it has functional significance (eg. transposon insertion target sites); nested - a repeat that is disrupted by the insertion of another element; dispersed - a repeat that is found dispersed throughout the genome; terminal - a repeat at the ends of and within the sequence for which it has functional significance (eg. transposon LTRs); long_terminal_repeat - a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses; non_ltr_retrotransposon_polymeric_tract - a polymeric tract, such as poly(dA), within a non LTR retrotransposon; centromeric_repeat - a repeat region found within the modular centromere; telomeric_repeat - a repeat region found within the telomere; x_element_combinatorial_repeat - a repeat region located between the X element and the telomere or adjacent Y' element; y_prime_element - a repeat region located adjacent to telomeric repeats or X element combinatorial repeats, either as a single copy or tandem repeat of two to four copies; other - a repeat exhibiting important attributes that cannot be described by other values.

6. 60.	Qual i fier	rpt_uni t_range
	Defi ni ti on	l oca ti on of a re pe a ti ng un i t ex pr e ss e d as a r a n g e
	Val ue fo r ma t	<ba se_r a n g e> - wh e r e <ba se_r a n g e> i s t h e f i r s t a n d l a s t b a s e (s e p a r a t e d b y t w o d o t s) o f a r e p e a t i n g u n i t
	Ex a m p l e	<I NSDQual i fi e r_val u e>202. . 245</I NSDQual i fi e r_val u e>
	Co m m e n t	u s e d t o i n d i c a t e t h e b a s e r a n g e o f t h e s e q u e n c e t h a t c o n s t i t u t e s a r e p e a t i n g u n i t w i t h i n t h e r e g i o n s p e c i f i e d b y t h e f e a t u r e k e y s o r i T a n d r e p e a t_r e g i o n.
6. 61.	Qual i fier	rpt_uni t_seq
	Defi ni ti on	i d e n t i t y o f a r e p e a t s e q u e n c e
	Val ue fo r ma t	f r e e t e x t
	Ex a m p l e	<I NSDQual i fi e r_val u e>aagggc</I NSDQual i fi e r_val u e> <I NSDQual i fi e r_val u e>ag(5)tg(8)</I NSDQual i fi e r_val u e> <I NSDQual i fi e r_val u e>(AAAGA)6(AAAA)1(AAAGA)12</I NSDQual i fi e r_val u e>
	Co m m e n t	u s e d t o i n d i c a t e t h e l i t e r a l s e q u e n c e t h a t c o n s t i t u t e s a r e p e a t i n g u n i t w i t h i n t h e r e g i o n s p e c i f i e d b y t h e f e a t u r e k e y s o r i T a n d r e p e a t_r e g i o n
6. 62.	Qual i fier	satellite
	Defi ni ti on	i d e n t i f i e r f o r a s a t e l l i t e D N A m a r k e r, c o m p o s e o f m a n y t a n d e m r e p e a t s (i d e n t i c a l o r r e l a t e d) o f a s h o r t b a s i c r e p e a t e d u n i t
	Val ue fo r ma t	<sa t e l l i t e_t y p e>[: <cl a s s>][<i d e n t i f i e r>] - wh e r e <sa t e l l i t e_t y p e> i s o n e o f t h e f o l l o w i n g: sa t e l l i t e; m i c r o s a t e l l i t e; m i n i s a t e l l i t e
	Ex a m p l e	<I NSDQual i fi e r_val u e>sa t e l l i t e: S1a</I NSDQual i fi e r_val u e> <I NSDQual i fi e r_val u e>sa t e l l i t e: a l p h a</I NSDQual i fi e r_val u e> <I NSDQual i fi e r_val u e>sa t e l l i t e: g a m m a I I I</I NSDQual i fi e r_val u e> <I NSDQual i fi e r_val u e>m i c r o s a t e l l i t e: D C 1 3 0</I NSDQual i fi e r_val u e>
	Co m m e n t	m a n y s a t e l l i t e s h a v e b a s e c o m p o s i t i o n o r o t h e r p r o p e r t i e s t h a t d i f f e r f r o m t h o s e o f t h e r e s t o f t h e g e n o m e t h a t a l l o w s t h e m t o b e i d e n t i f i e d.
6. 63.	Qual i fier	segment
	Defi ni ti on	n a m e o f v i r a l o r p h a g e s e g m e n t s e q u e n c e d
	Val ue fo r ma t	f r e e t e x t
	Ex a m p l e	<I NSDQual i fi e r_val u e>6</I NSDQual i fi e r_val u e>
6. 64.	Qual i fier	serotype
	Defi ni ti on	s e r o l o g i c a l v a r i e t y o f a s p e c i e s c h a r a c t e r i z e d b y i t s a n t i g e n i c p r o p e r t i e s
	Val ue fo r ma t	f r e e t e x t
	Ex a m p l e	<I NSDQual i fi e r_val u e>B1</I NSDQual i fi e r_val u e>
	Co m m e n t	u s e d o n l y w i t h t h e s o u r c e f e a t u r e k e y; t h e B a c t e r i o l o g i c a l C o d e r e c o m m e n d s t h e u s e o f t h e t e r m 's e r o v a r' i n s t e a d o f 's e r o t y p e' f o r t h e p r o k a r y o t e s; s e e t h e I n t e r n a t i o n a l C o d e o f N o m e n c l a t u r e o f B a c t e r i a (1990 R e v i s i o n) A p p e n d i x 10. B "I n f r a s p e c i f i c T e r m s".

6. 65.	Qual i fi er	serovar
	Defi ni ti on	serol ogical variety of a species (usual ly a prokaryote) characterized by its antigenic propert ies
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>0157: H7</INSDQual i fi er_val ue>
	Comment	used only wi th the source feature key; the Bacteri ological Code recommends the use of the term 'serovar' instead of 'serotype' for prokaryotes; see the International Code of Nomenclature of Bacteri a (1990 Revisi on) Appendi x 10. B "Infraspeci fic Terms".
6. 66.	Qual i fi er	sex
	Defi ni ti on	sex of the organ ism from which the sequence was obtained; sex is used for eukaryotic organ isms that undergo mei osi s and have sexual ly di morphic gametes
	Val ue format	free text
	Exampl es	<INSDQual i fi er_val ue>femal e</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>mal e</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>hermaphrodi te</INSDQual i fi er_val ue> <INSDQual i fi er_val ue>uni sexual </INSDQual i fi er_val ue> <INSDQual i fi er_val ue>bi sexual </INSDQual i fi er_val ue> <INSDQual i fi er_val ue>asexual </INSDQual i fi er_val ue> <INSDQual i fi er_val ue>monoeci ous</INSDQual i fi er_val ue> [or moneci ous] <INSDQual i fi er_val ue>di oeci ous</INSDQual i fi er_val ue> [or di eci ous]
	Comment	The sex qual i fi er should be used (instead of mating_type qual i fi er) in the Metazoa, Embryophyta, Rhodophyta & Phaeophyceae; mating_type qual i fi er should be used (instead of sex qual i fi er) in the Bacteri a, Archaea & Fungi; neither sex nor mating_type qual i fi ers should be used in the viruses; outside of the taxa listed above, mating_type qual i fi er should be used unless the value of the qual i fi er is taken from the vocabulary given in the exampl es above
6. 67.	Qual i fi er	standard_name
	Defi ni ti on	accepted standard name for this feature
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>dotted</INSDQual i fi er_val ue>
	Comment	use standard_name qual i fi er to give full gene name, but use gene qual i fi er to give gene symbol (in the above exampl e gene qual i fi er value is Dt).
6. 68.	Qual i fi er	strain
	Defi ni ti on	strain from which sequence was obtained
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>BALB/c</INSDQual i fi er_val ue>
	Comment	entries including strain qual i fi er must not include the environmental_sample qual i fi er

6. 69.	Qual i fi er	sub_cl one
	Defi ni ti on	sub-cl one from whi ch sequence was obtai ned
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>l ambda-hl L7. 20g</INSDQual i fi er_val ue>
	Comment	not more than one sub_cl one should be speci fi ed for the source feature; to indi cate that the sequence was obtai ned from mul ti ple sub_cl ones, mul ti ple sources may be further descri bed using the feature key "mi sc_feature" and the qual i fi er "note"
6. 70.	Qual i fi er	sub_speci es
	Defi ni ti on	name of sub-speci es of organi sm from whi ch sequence was obtai ned
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>l acti s</INSDQual i fi er_val ue>
6. 71.	Qual i fi er	sub_strain
	Defi ni ti on	name or i denti fi er of a geneti cally or otherwi se modi fi ed strain from whi ch sequence was obtai ned, derived from a parenta l strain (whi ch should be annota ted in the strain qual i fi er). sub_strain from whi ch sequence was obtai ned
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>abi s</INSDQual i fi er_val ue>
	Comment	If the parenta l strain is not given, thi s should be annota ted in the strain qual i fi er instea d of sub_strain. For exampl e, ei ther a strain qual i fi er wi th the val ue K-12 and a substra in qual i fi er wi th the val ue MG1655 or a strain qual i fi er wi th the val ue MG1655
6. 72.	Qual i fi er	tag_pepti de
	Defi ni ti on	base loca ti on encodi ng the polypepti de for proteolysi s tag of tmRNA and i ts termina ti on codon
	Val ue format	<base_range> - where <base_range> provi des the fi rst and last base (separa ted by two dots) of the loca ti on for the proteolysi s tag
	Exampl e	<INSDQual i fi er_val ue>90. . 122</INSDQual i fi er_val ue>
	Comment	i t is recomme nded that the ami no aci d sequence correspon di ng to the tag_pepti de be annota ted by descri bi ng a 5' parti al CDS feature; e. g. CDS wi th a loca ti on of <90. . 122
6. 73.	Qual i fi er	tissu e_li b
	Defi ni ti on	tissu e li brary from whi ch sequence was obtai ned
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>tissu e li brary 772</INSDQual i fi er_val ue>
6. 74.	Qual i fi er	tissu e_type
	Defi ni ti on	tissu e type from whi ch the sequence was obtai ned
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>l i ver</INSDQual i fi er_val ue>

6.75.	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 three letter abbreviation for the amino acid coded by the codon at the base_range position
	Example	<INSDQualifier_value>(pos: 213..215, aa: Trp) </INSDQualifier_value> <INSDQualifier_value>(pos: 462..464, aa: OTHER) </INSDQualifier_value> <INSDQualifier_value>(pos: 1017, aa: TERM) </INSDQualifier_value> <INSDQualifier_value>(pos: 2000..2001, aa: TERM) </INSDQualifier_value>
	Comment	if the amino acid is not one of the specific amino acids listed in Section 3 of this Annex, use OTHER as <amino_acid> and provide the name of the unusual amino acid in a note qualifier; for modified amino-acid selenocysteine use three letter abbreviation 'Sec' (one letter symbol 'U' in amino-acid sequence) for <amino_acid>; for modified amino-acid pyrrolysine use three letter abbreviation 'Pyl' (one letter symbol 'O' in amino-acid sequence) for <amino_acid>; 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 for the location, see the third and fourth examples above, in conjunction with a note qualifier indicating 'stop codon completed by the addition of 3' A residues to the mRNA'.
6.76.	Qualifier	transl_table
	Definition	definition of genetic code table used if other than universal or standard genetic code table. Tables used are described in this Annex
	Value format	<integer> where <integer> is the number assigned to the genetic code table
	Example	<INSDQualifier_value>3</INSDQualifier_value> - example where the yeast mitochondrial code is to be used
	Comment	if the transl_table qualifier is not used to further annotate a CDS feature key, then the CDS is translated using the Standard Code (i.e. Universal Genetic Code). Genetic code exceptions outside the range of specified tables are reported in transl_except qualifiers.
6.77.	Qualifier	trans_splicing
	Definition	indicates that exons from two RNA molecules are ligated in intermolecular reaction to form mature RNA
	Value format	none
	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, e.g. join(complement(69611..69724), 139856..140087)
6.78.	Qualifier	translation
	Definition	one-letter abbreviated amino acid sequence derived from either the standard (or universal) genetic code or the table as specified in a transl_table qualifier and as determined by an exception in the transl_except qualifier
	Value format	contiguous string of one-letter amino acid abbreviations from Section 3 of this Annex, "X" is to be used for AA exceptions.
	Example	<INSDQualifier_value>MASTFPWPYRGCASTPSLKGLI MCTW</INSDQualifier_value>
	Comment	to be used with CDS feature only; must be accompanied by protein_id qualifier when the translation product contains four or more specifically defined amino acids; see transl_tables for definition and location of genetic code tables; only one of the qualifiers translation, pseudo and pseudogene are permitted to further annotate a

CDS feature.

6.79.	Qual i fi er	vari ety
	Defi ni ti on	vari ety (= vari etas, a formal Linnaean rank) of organi sm from whi ch sequence was derived.
	Val ue format	free text
	Exampl e	<INSDQual i fi er_val ue>i nsul ari s</INSDQual i fi er_val ue>
	Comment	use the cultivar qual i fi er for cultivated plant varieties, i .e., products of arti fi cial selection; varieties other than plant and fungal vari etas should be annotated via a note qual i fi er, e.g. wi th the value <INSDQual i fi er_val ue>breed: Cukorova</INSDQual i fi er_val ue>

SECTION 7: FEATURE KEYS FOR AMINO ACID SEQUENCES

This section contains the list of allowed feature keys to be used for amino acid sequences. The feature keys are listed in alphabetic order.

7.1.	Feature Key	ACT_SITE
	Definition	Amino acid(s) involved in the activity of an enzyme
	Optional qualifiers	NOTE
	Comment	Each amino acid residue of the active site should be annotated separately with the ACT_SITE feature key. The corresponding amino acid residue number should be provided as the location descriptor in the feature location element.
7.2.	Feature Key	BINDING
	Definition	Binding site for any chemical group (co-enzyme, prosthetic group, etc.). The chemical nature of the group is indicated in the NOTE qualifier
	Mandatory qualifiers	NOTE
	Comment	Examples of values for the "NOTE" qualifier: "Heme (covalent)" and "Chloride." Where appropriate, the features keys CA_BIND, DNA_BIND, METAL, and NP_BIND should be used rather than BINDING.
7.3.	Feature Key	CA_BIND
	Definition	Extent of a calcium-binding region
	Optional qualifiers	NOTE
7.4.	Feature Key	CARBOHYD
	Definition	Glycosylation site
	Mandatory qualifiers	NOTE
	Comment	This key describes the occurrence of the attachment of a glycan (mono- or polysaccharide) to a residue of the protein. The type of linkage (C-, N- or O-linked) to the protein is indicated in the "NOTE" qualifier. If the nature of the reducing terminal sugar is known, its abbreviation is shown between parentheses. If three dots '...' follow the abbreviation this indicates an extension of the carbohydrate chain. Conversely no dots means that a monosaccharide is linked. Examples of values used in the "NOTE" qualifier: N-linked (GlcNAc...); O-linked (GlcNAc); O-linked (Glc...); C-linked (Man) partial; O-linked (Ara...).

7.5.	Feature Key	CHAIN
	Definition	Extent of a polypeptide chain in the mature protein
	Optional qualifiers	NOTE
7.6.	Feature Key	COILED
	Definition	Extent of a coiled-coil region
	Optional qualifiers	NOTE
7.7.	Feature Key	COMPBIAS
	Definition	Extent of a compositionally biased region
	Optional qualifiers	NOTE
7.8.	Feature Key	CONFLICT
	Definition	Different sources report differing sequences
	Optional qualifiers	NOTE
	Comment	Examples of values for the "NOTE" qualifier: Missing; K -> Q; GSDSE -> RIRLR; V -> A.
7.9.	Feature Key	CROSSLNK
	Definition	Post translationally formed amino acid bonds
	Mandatory qualifiers	NOTE
	Comment	Covalent linkages of various types formed between two proteins (interchain cross-links) or between two parts of the same protein (intrachain cross-links); except for cross-links formed by disulfide bonds, for which the "DISULFID" feature key is to be used. For an interchain cross-link, the location descriptor in the feature location element is the residue number of the amino acid cross-linked to the other protein. For an intrachain cross-link, the location descriptors in the feature location element are the residue numbers of the cross-linked amino acids in conjunction with the "join" location operator, e.g. "join(42,50)". The NOTE qualifier indicates the nature of the cross-link; at least specifying the name of the conjugate and the identity of the two amino acids involved. Examples of values for the "NOTE" qualifier: "Isoglutamyl cysteine thioester (Cys-Gln);" "Beta-methylanthionine (Cys-Thr);" and "Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin)"
7.10.	Feature Key	DISULFID
	Definition	Disulfide bond
	Optional qualifiers	NOTE
	Comment	For an interchain disulfide bond, the location descriptor in the feature location element is the residue number of the cysteine linked to the other protein. For an intrachain cross-link, the location descriptors in the feature location element are the residue numbers of the linked cysteines in conjunction with the "join" location operator, e.g. "join(42,50)". For interchain disulfide bonds, the NOTE qualifier indicates the nature of the cross-link, by identifying the other protein, for example, "Interchain (between A and B chains)"

7.11.	Feature Key	DNA_BIND
	Definition	Extent of a DNA-binding region
	Mandatory qualifiers	NOTE
	Comment	The nature of the DNA-binding region is given in the NOTE qualifier. Examples of values for the "NOTE" qualifier: "Homeobox" and "Myb 2"
7.12.	Feature Key	DOMAIN
	Definition	Extent of a domain, which is defined as a specific combination of secondary structures organized into a characteristic three-dimensional structure or fold
	Mandatory qualifiers	NOTE
	Comment	The domain type is given in the NOTE qualifier. Where several copies of a domain are present, the domains are numbered. Examples of values for the "NOTE" qualifier: "Ras-GAP" and "Cadherin 1"
7.13.	Feature Key	HELIX
	Definition	Secondary structure: Helices, for example, Alpha-helix; 3(10) helix; or Pi-helix
	Optional qualifiers	NOTE
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.
7.14.	Feature Key	INIT_MET
	Definition	Initiator methionine
	Optional qualifiers	NOTE
	Comment	The location descriptor in the feature location element is "1". This feature key indicates the N-terminal methionine is cleaved off. This feature is not used when the initiator methionine is not cleaved off.
7.15.	Feature Key	INTRAMEM
	Definition	Extent of a region located in a membrane without crossing it
	Optional qualifiers	NOTE
7.16.	Feature Key	LIPID
	Definition	Covalent binding of a lipid moiety
	Mandatory qualifiers	NOTE
	Comment	The chemical nature of the bound lipid moiety is given in the NOTE qualifier, indicating at least the name of the lipidated amino acid. Examples of values for the "NOTE" qualifier: "N-myristoyl glycine"; "GPI-anchor amidated serine" and "S-diacylglycerol cysteine."

7.17.	Feature Key	METAL
	Defini tion	Binding site for a metal ion.
	Mandatory qual i fiers	NOTE
	Comment	The NOTE qualifier indicates the nature of the metal. Examples of values for the "NOTE" qualifier: "Iron (heme axial ligand)" and "Copper".
7.18.	Feature Key	MOD_RES
	Defini tion	Posttranslational modi fication of a residue
	Mandatory qual i fiers	NOTE
	Comment	The chemical nature of the modified residue is given in the NOTE qualifier, indicating at least the name of the post-translationally modified amino acid. If the modified amino acid is listed in Section 4 of this Annex, the abbreviation may be used in place of the the full name. Examples of values for the "NOTE" qualifier: "N-acetylal anine"; "3-Hyp"; and "MeLys" or "N-6-methyl lysine"
7.19.	Feature Key	MOTIF
	Defini tion	Short (up to 20 amino acids) sequence motif of biological interest
	Optional qual i fiers	NOTE
7.20.	Feature Key	MUTAGEN
	Defini tion	Site which has been experimentally altered by mutagenesis
	Optional qual i fiers	NOTE
7.21.	Feature Key	NON_STD
	Defini tion	Non-standard amino acid
	Optional qual i fiers	NOTE
	Comment	This key describes the occurrence of non-standard amino acids selenocysteine (U) and pyrrolysine (O) in the amino acid sequence.
7.22.	Feature Key	NON_TER
	Defini tion	The residue at an extremity of the sequence is not the terminal residue
	Optional qual i fiers	NOTE
	Comment	If applied to position 1, this means that the first position is not the N-terminus of the complete molecule. If applied to the last position, it means that this position is not the C-terminus of the complete molecule.
7.23.	Feature Key	NP_BIND
	Defini tion	Extent of a nucleotide phosphate-binding region
	Mandatory qual i fiers	NOTE
	Comment	The nature of the nucleotide phosphate is indicated in the NOTE qualifier. Examples of values for the "NOTE" qualifier: "ATP" and "FAD".

7.24.	Feature Key	PEPTIDE
	Definition	Extent of a released active peptide
	Optional qualifiers	NOTE
7.25.	Feature Key	PROPEP
	Definition	Extent of a propeptide
	Optional qualifiers	NOTE
7.26.	Feature Key	REGION
	Definition	Extent of a region of interest in the sequence
	Optional qualifiers	NOTE
7.27.	Feature Key	REPEAT
	Definition	Extent of an internal sequence repetition
	Optional qualifiers	NOTE
7.28.	Feature Key	SIGNAL
	Definition	Extent of a signal sequence (prepeptide)
	Optional qualifiers	NOTE
7.29.	Feature Key	SITE
	Definition	Any interesting single amino-acid site on the sequence that is not defined by another feature key. It can also apply to an amino acid bond which is represented by the positions of the two flanking amino acids
	Mandatory qualifier	NOTE
	Comment	When SITE is used to annotate a modified amino acid the value for the qualifier "NOTE" must either be an abbreviation set forth in Section 4 of this Annex, or the complete, unabbreviated name of the modified amino acid.
7.30.	Feature Key	SOURCE
	Definition	Identifies the source of the sequence; this key is mandatory; every sequence will have a single SOURCE feature spanning the entire sequence
	Mandatory qualifiers	MOL_TYPE ORGANISM
	Optional qualifiers	NOTE

7.31.	Feature Key	STRAND
	Defini tion	Secondary structure: Beta-strand; for example Hydrogen bonded beta-strand or residue in an isolated beta-bridge
	Optional qual ifiers	NOTE
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are speci fied: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not speci fied in one of these classes are in a 'loop' or 'random-coil' structure.
7.32.	Feature Key	TOPO_DOM
	Defini tion	Topol ogical domai n
	Optional qual ifiers	NOTE
7.33.	Feature Key	TRANSMEM
	Defini tion	Extent of a transmembrane regi on
	Optional qual ifiers	NOTE
7.34.	Feature Key	TRANSIT
	Defini tion	Extent of a transi t pepti de (mi tochondri on, chl oroplast, thylakoi d, cyanelle, peroxisome etc.)
	Optional qual ifiers	NOTE
7.35.	Feature Key	TURN
	Defini tion	Secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)
	Optional qual ifiers	NOTE
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are speci fied: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not speci fied in one of these classes are in a 'loop' or 'random-coil' structure.
7.36.	Feature Key	UNSURE
	Defini tion	Uncertainties in the sequence
	Optional qual ifiers	NOTE
	Comment	Used to describe regi on(s) of an ami no acid sequence for which the authors are unsure about the sequence presentation.
7.37.	Feature Key	VARIANT
	Defini tion	Authors report that sequence vari ants exist
	Optional qual ifiers	NOTE

7.38.	Feature Key	VAR_SEQ
	Definition	Description of sequence variants produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting
	Optional qualifiers	NOTE

7.39.	Feature Key	ZN_FING
	Definition	Extent of a zinc finger region
	Mandatory qualifiers	NOTE
	Comment	The type of zinc finger is indicated in the NOTE qualifier. For example: "GATA-type" and "NR C4-type"

SECTION 8: QUALIFIERS FOR AMINO ACID SEQUENCES

This section contains the list of allowed qualifiers to be used for amino acid sequences.

8.1.	Qualifier	MOL_TYPE
	Definition	In vivo molecule type of sequence
	Value format	protein
	Example	<INSDQualifier_val ue>protei n</INSDQualifier_val ue>
	Comment	The "MOL_TYPE" qualifier is mandatory on the SOURCE feature key.
8.2.	Qualifier	NOTE
	Definition	Any comment or additional information
	Value format	free text
	Example	<INSDQualifier_val ue>Heme (coval ent)</INSDQualifier_val ue>
	Comment	The "NOTE" qualifier is mandatory for the feature keys: BINDING; CARBOHYD; CROSSLNK; DI SULFID; DNA_BIND; DOMAIN; LIPID; METAL; MOD_RES; NP_BIND and ZN_FING
8.3.	Qualifier	ORGANISM
	Definition	Scientific name of the organism that provided the peptide
	Value format	free text
	Example	<INSDQualifier_val ue>Homo sapi ens</INSDQualifier_val ue>
	Comment	The "ORGANISM" qualifier is mandatory for the SOURCE feature key.

