

Приложение 1

Контролируемая лексика

Версия 1.6

*Пересмотр одобрен Комитетом по Стандартам ВОИС (КСВ)
на своей десятой сессии 25 ноября 2022*

Содержание

Раздел 1 Перечень нуклеотидов	2
Раздел 2 Перечень модифицированных нуклеотидов.....	2
Раздел 3 Перечень аминокислот.....	5
Раздел 4 Перечень модифицированных аминокислот	6
Раздел 5 Функциональные ключи для нуклеотидных последовательностей.....	7
Раздел 6 Квалификаторы для нуклеотидных последовательностей	25
Раздел 7 Функциональные ключи для аминокислотных последовательностей	51
Раздел 8 Квалификаторы для аминокислотных последовательностей	59
Раздел 9 Таблицы генетических кодов	61

Раздел 1 Перечень нуклеотидов

Условные обозначения основных нуклеотидов, которые используются в перечне последовательностей, представлены в Таблице 1. Символ “t” будет истолкован как тимин в ДНК и урацил в РНК, если он используется без последующего описания. В случае, когда присвоен неоднозначный символ (представляющий два и более альтернативных основания), должен быть использован наиболее ограничивающий символ. Например, в случае если основаниями данной позиции могут быть “a или g”, следует использовать скорее “r”, чем “n”. Символ “n” будет рассматриваться как “a или c или g или t/u”, если он используется без дополнительного описания.

Таблица 1: Перечень обозначений нуклеотидов

Символ	Определение
a	аденин /adenine
c	цитозин /cytosine
g	гуанин /guanine
t	тимин в ДНК/урацил в РНК (t/u)thymine in DNA/uracil in RNA (t/u)
m	a или c
r	a или g
w	a или t/u
s	c или g
y	c или t/u
k	g или t/u
v	a или c или g; не t/u
h	a или c или t/u; не g
d	a или g или t/u; не c
b	c или g или t/u; не a
n	a или c или g или t/u; неизвестная или другая

Раздел 2 Перечень модифицированных нуклеотидов

Аббревиатуры, перечисленные в Таблице 2, являются единственными разрешёнными значениями для квалификатора mod_base. Если определённый модифицированный нуклеотид не представлен в нижеприведённой таблице, тогда в качестве его значения должна использоваться аббревиатура “OTHER”. Если аббревиатурой является “OTHER”, то полное несокращённое название модифицированного основания должно быть приведено в квалификаторе “note”. Аббревиатуры, приведенные в Таблице 2, не должны быть использованы в самой последовательности.

Таблица 2: Перечень модифицированных нуклеотидов

Аббревиатура	Определение
ac4c	4-ацетилцитидин 4-acetylcytidine
chm5u	5-(карбоксигидроксиметил)уридин 5-(carboxyhydroxymethyl)uridine
cm	2'-О-метилцитидин 2'-O-methylcytidine

Аббревиатура	Определение
cmnm5s2u	5-карбоксиметиламинометил-2-тиоуридин (5-carboxymethylaminomethyl-2-thiouridine)
cmnm5u	5-карбоксиметиламинометилуридин (5-carboxymethylaminomethyluridine)
dhu	дигидроуридин (dihydrouridine)
fm	2'-О-метилпсевдоуридин (2'-O-methylpseudouridine)
gal q	бета,D-галактозилквеуозин (beta,D-galactosylqueuosine)
gm	2'-О-метилгуанозин (2'-O-methylguanosine)
i	ИНОЗИН (inosine)
i6a	N6-изопениладенозин (N6-isopentenyladenosine)
m1a	1-метиладенозин (1-methyladenosine)
m1f	1-метилпсевдоуридин (1-methylpseudouridine)
m1g	1-метилгуанозин (1-methylguanosine)
m1i	1-метиинозин (1-methylinosine)
m22g	2,2-диметилгуанозин (2,2-dimethylguanosine)
m2a	2-метиладенозин (2-methyladenosine)
m2g	2-метилгуанозин (2-methylguanosine)
m3c	3-метилцитидин (3-methylcytidine)
m4c	N4-methylcytosine (N4-methylcytosine)
m5c	5-метилцитидин (5-methylcytidine)
m6a	N6-метиладенозин (N6-methyladenosine)
m7g	7-метилгуанозин (7-methylguanosine)
mam5u	5-метиламинометилуридин (5-methylaminomethyluridine)
mam5s2u	5- метоксиаминометил-2-тиоуридин (5-methoxyaminomethyl-2-thiouridine)
man q	бета,D-манозилквеуозин (beta,D-mannosylqueuosine)

Аббревиатура	Определение
mcm5s2u	5-метиламинометил-2-тиоуридин (5-methylaminomethyl-2-thiouridine)
mcm5u	5-метоксикарбонилметилуридин (5-methoxycarbonylmethyluridine)
mo5u	5-метоксиуридин (5-methoxyuridine)
ms2i6a	2-метилтио-N6-изопентиладенозин (2-methylthio-N6-isopentenyladenosine)
ms2t6a	N-((9-бета-D-рибофуранозил-2-метилтиопурин-6-ил)карбамойл)треонин (N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine)
mt6a	N-((9-бета-D-рибофуранозилпурин-6-ил)N-метилкарбамойл)треонин (N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine)
mv	уридин-5-оксоуксусная кислота-метилэфир (uridine-5-oxoacetic acid-methylester)
o5u	уридин-5-оксоуксусная кислота (v) (uridine-5-oxoacetic acid (v))
osyw	вибутоксозин (wybutoxosine)
p	псевдоуридин (pseudouridine)
q	квеуозин (queuosine)
s2c	2-тиоцитидин (2-thiocytidine)
s2t	5-метил-2-тиоуридин (5-methyl-2-thiouridine)
s2u	2-тиоуридин (2-thiouridine_
s4u	4-тиоуридин (4-thiouridine)
m5u	5-метилуридин (5-methyluridine)
t6a	N-((9-бета-D-рибофураназилпурин-6-ил)карбамойл)треонин (N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine)
tm	2'-О-метил-5-метилуридин (2'-O-methyl-5-methyluridine)
um	2'-О-метилуридин (2'-O-methyluridine)
yw	вибутозин (wybutosine)
x	3-(3-амино-3-карбоксипропил)уридин, (аср3)u (3-(3-amino-3-carboxypropyl)uridine, (аср3)u)
OTHER	(требуется квалификатор "note")

Раздел 3 Перечень аминокислот

Условные обозначения аминокислот, которые используются в последовательностях, представлены в Таблице 3. Когда присвоен неоднозначный символ (представляющий две или более альтернативных аминокислоты), должен быть использован наиболее ограничивающий символ. Например, если аминокислота в данной позиции может быть аспарагиновой кислотой или аспарагином, следует использовать скорее символ “B”, чем “X”. Символ “X” будет рассматриваться как один из символов “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, или “V”, если он используется без дополнительного описания.

Таблица 3: Перечень обозначений аминокислот

Символ	Определение
A	Аланин/Alanine
R	Аргинин/Arginine
N	Аспарагин/Asparagine
D	Аспарагиновая кислота (Аспартат)/ Aspartic acid (Aspartate)
C	Цистеин /Cysteine
Q	Глутамин /Glutamine
E	Глутаминовая кислота (Глутамат)/ Glutamic acid (Glutamate)
G	Глицин/Glycine
H	Гистидин/Histidine
I	Изолейцин/Isoleucine
L	Лейцин/Leucine
K	Лизин/Lysine
M	Метионин/Methionine
F	Фенилаланин/Phenylalanine
P	Пролин/Proline
O	Пирролизин/Pyrrolysine
S	Серин/Serine
U	Селеноцистеин/Selenocysteine
T	Треонин/Threonine
W	Триптофан/Tryptophan
Y	Тирозин/Tyrosine
V	Валин/Valine
B	Аспарагиновая кислота или Аспарагин/Aspartic acid or Asparagine
Z	Глутамин или Глутаминовая кислота/Glutamine or Glutamic acid
J	Лейцин или Изолейцин/Leucine or Isoleucine
X	A или R или N или D или C или Q или E или G или H или I или L или K или M или F или P или O или S или U или T или W или Y или V; “unknown” или “other”

Раздел 4 Перечень модифицированных аминокислот

В Таблице 4 перечислены единственные разрешённые аббревиатуры для модифицированных аминокислот в обязательном квалификаторе “note” для функциональных ключей “MOD_RES” или “SITE”. Значение для квалификатора “note” должно быть или одной из аббревиатур из этой таблицы, где это возможно, или полным, несокращённым названием модифицированной аминокислоты. Аббревиатуры (или полные названия), приведённые в данной таблице, не должны быть использованы в самой последовательности.

Таблица 4: 4 Перечень модифицированных аминокислот

Аббревиатура	Модифицированная аминокислота
Aad	2-аминоадипиновая кислота/2-Aminoadipic acid
bAad	3-аминоадипиновая кислота/3-Aminoadipic acid
bAla	бета-аланин, бета-аминопропионовая кислота/ beta-Alanine, beta-Aminopropionic acid
Abu	2-аминобутириновая кислота/2-Aminobutyric acid
4Abu	4-аминобутириновая кислота, пиперидиновая кислота/ 4-Aminobutyric acid, piperidinic acid
Acpr	6-аминокапроиновая кислота/6-Aminocaproic acid
Ahe	2-аминогептаноиновая кислота/2-Aminoheptanoic acid
Aib	2-аминоизобутириновая кислота/2-Aminoisobutyric acid
bAib	3-аминоизобутириновая кислота/3-Aminoisobutyric acid
Apm	2-аминопимелиновая кислота/2-Aminopimelic acid
Dbu	2,4-диаминобутириновая кислота /2,4-Diaminobutyric acid
Des	десмозин/Desmosine
Dpm	2,2'-диаминопимелиновая кислота/2,2'-Diaminopimelic acid
Dpr	2,3-диаминопропионовая кислота/2,3-Diaminopropionic acid
EtGly	N-этилглицин/N-Ethylglycine
EtAsn	N-этиласпарагин/N-Ethylasparagine
Hyl	гидроксилизин/Hydroxylysine
aHyl	алло-гидроксилизин/allo-Hydroxylysine
3Hyp	3-гидроксипролин/3-Hydroxyproline
4Hyp	4-гидроксипролин/4-Hydroxyproline
Ide	Изодезмосин/Isodesmosine
aIle	алло-изолейцин/allo-Isoleucine
MeGly	N-метилглицин, саркозин/N-Methylglycine, sarcosine
MeIle	N- метилизолейцин/N-Methylisoleucine
MeLys	6-N-метиллизин/6-N-Methyllysine
MeVal	N-метилвалин/N-Methylvaline
Nva	Норвалин/Norvaline
Nle	Норлейцин/Norleucine
Orn	Орнитин/Ornithine

Раздел 5 Функциональные ключи для нуклеотидных последовательностей

Данная глава содержит перечень разрешённых функциональных ключей для использования в нуклеотидных последовательностях, и перечень обязательных и необязательных квалификаторов. Функциональные ключи перечислены в алфавитном порядке. Функциональные ключи могут быть использованы для ДНК или для РНК, если другое не указано в поле “Molecule scope”. Некоторые функциональные ключи могут подходить для использования с не природными последовательностями в дополнение к определенным “organism scope”.

Названия функциональных ключей в случае XML последовательности должны использоваться в точности так, как они приводятся в описаниях ниже в поле “Feature key”, за исключением функциональных ключей 3'UTR и 5'UTR. Для функциональных ключей 3'UTR и 5'UTR см. поле “Comment” в описании.

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 circular_RNA 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, pseudogene or 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
Definition	region of biological interest identified as a centromere and which has been experimentally characterized
Optional qualifiers	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
Definition	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
Optional qualifiers	allele gene gene_synonym map note
Molecule scope	DNA
5.5.Feature Key	D_segment
Definition	Diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
Optional qualifiers	allele gene gene_synonym

	map note product pseudo pseudogene standard_name
Organism scope	eukaryotes
5.6.Feature Key	exon
Definition	region of genome that codes for portion of spliced mRNA, rRNA and tRNA; may contain 5'UTR, all CDSs and 3' UTR
Optional qualifiers	allele EC_number function gene gene_synonym map note number product pseudo pseudogene standard_name trans_splicing
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
Parent Key	CDS
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_classqualifier with the value "ribosome_binding_site" must be used for describing ribosome binding sites
5.13 Feature Key	misc_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 misc_difference feature key must 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. The variation feature key must be used to describe naturally occurring genetic variability.

5.14 Feature Key	misc_feature
Definition	region of biological interest which cannot be described by any other feature key; a new or rare feature
Optional qualifiers	allele 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	misc_recomb
Definition	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys or qualifiers of source key (proviral)
Optional qualifiers	allele gene gene_synonym map note recombination_class standard_name
Molecule scope	DNA

5.16 Feature Key	misc_RNA
Definition	any transcript or RNA product that cannot be defined by other RNA keys (prim_transcript, precursor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, ncRNA, rRNA and tRNA)
Optional qualifiers	allele function gene gene_synonym map note

	operon product pseudo pseudogene standard_name trans_splicing
5.17 Feature Key	misc_structure
Definition	any secondary or tertiary nucleotide structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
Optional qualifiers	allele function gene gene_synonym map note standard_name
5.18 Feature Key	mobile_element
Definition	region of genome containing mobile elements
Mandatory qualifiers	mobile_element_type
Optional qualifiers	allele function gene gene_synonym map note rpt_family rpt_type standard_name
5.19 Feature Key	modified_base
Definition	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
Mandatory qualifiers	mod_base
Optional qualifiers	allele frequency gene gene_synonym map note
Comment	value for the mandatory mod_base qualifier is limited to the restricted vocabulary for modified 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 circular_RNA 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 must not be used for ribosomal and transfer RNA annotation, for which the rRNA and tRNA feature keys must 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

Organism scope	pseudogene standard_name eukaryotes
5.23. Feature Key	operon
Definition	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
Mandatory qualifiers	operon
Optional qualifiers	allele function map note phenotype pseudo pseudogene standard_name
5.24. Feature Key	oriT
Definition	origin of transfer; region of a DNA molecule where transfer is initiated during the process of conjugation or mobilization
Optional qualifiers	allele bound_moiety direction gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq standard_name
Molecule scope	DNA
Comment	rep_origin must be used to describe origins of replication; direction qualifier has legal values left, right and both, however only left and right are valid when used in conjunction with the oriT feature; origins of transfer can be present in the chromosome; plasmids can contain multiple origins of transfer
5.25 Feature Key	polyA_site
Definition	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation

Optional qualifiers	allele gene gene_synonym map note
Organism scope	eukaryotes and eukaryotic viruses

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	propeptide
Definition	propeptide coding sequence; coding sequence for the domain of a proprotein that is cleaved to form the mature protein product.
Optional qualifiers	allele function gene gene_synonym map note product pseudo pseudogene standard_name
5.30. 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" must be used to describe ribosome binding sites
5.31. Feature Key	regulatory
Definition	any region of a sequence that functions in the regulation of transcription, translation, replication or chromatin structure;
Mandatory qualifiers	regulatory_class
Optional qualifiers	allele bound_moiety

	function
	gene
	gene_synonym
	map
	note
	operon
	phenotype
	pseudo
	pseudogene
	standard_name
5.32. 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
	standard_name
5.33. Feature Key	rep_origin
Definition	origin of replication; starting site for duplication of nucleic acid to give two identical copies
Optional Qualifiers	allele
	direction
	function
	gene
	gene_synonym
	map
	note
	standard_name
Comment	direction qualifier has valid values: left, right, or both
5.34. 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
5.35. 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
5.36. 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
5.37. 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 proviral rearranged segment serotype serovar sex strain sub_clone sub_species sub_strain tissue_lib tissue_type variety
Molecule scope	any
5.38. 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	
5.39. 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
5.40. 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
5.41. 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	
5.42. 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
5.43. 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 circular_RNA anticodon function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.44. 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.

5.45. 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

5.46. 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

5.47. 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; use the replace qualifier to annotate a deletion, insertion, or substitution; variability arising as a result of genetic manipulation (e.g. site directed mutagenesis) must be described with the misc_difference feature;
5.48. 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>.
5.49. 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>.

Раздел 6 Квалификаторы для нуклеотидных последовательностей

Данная глава содержит перечень квалификаторов, которые будут использоваться для описания признаков в нуклеотидных последовательностях. Квалификаторы перечислены в алфавитном порядке.

Если формат значения “none”, элементы INSDQualifier_value и NonEnglishQualifier_value не должны использоваться.

Если формат значения представляет собой свободный текст, который определяется как зависящий от языка, необходимо использовать одно из следующих значений:

- 1) элемент INSDQualifier_value; или же
- 2) элемент NonEnglishQualifier_value; или же
- 3) как элемент INSDQualifier_value, так и элемент NonEnglishQualifier_value.

ПОЖАЛУЙСТА, ОБРАТИТЕ ВНИМАНИЕ: Любое значение квалификатора, предусмотренное для квалификатора со значением формата «свободный текст» (“free text”) зависящий от языка, может потребовать перевода для международных, национальных или региональных процедур.

Квалификаторы, перечисленные в следующей таблице, имеют значения свободного текста зависящего от языка:

Таблица 5: Список значений квалификаторов для нуклеотидных последовательностей с зависимыми от языка значениями свободного текста

Часть	Квалификатор свободного текста зависящего от языка
6.3	bound_moiety
6.5	cell_type
6.8	clone
6.9	clone_lib
6.11	collected_by
6.14	cultivar
6.15	dev_stage
6.18	ecotype
6.21	frequency
6.22	function
6.24	gene_synonym
6.26	haplogroup
6.28	host
6.29	identified_by
6.30	isolate
6.31	isolation_source
6.32	lab_host
6.36	mating_type
6.41	note
6.45	organism
6.47	phenotype
6.49	pop_variant
6.50	product
6.66	serotype
6.67	serovar
6.68	sex
6.69	standard_name
6.70	strain
6.71	sub_clone
6.72	sub_species
6.73	sub_strain
6.75	tissue_lib
6.76	tissue_type
6.81	variety

6.1.	Qualifier	allele
Definition	name of the allele for the given gene	
Mandatory 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	
Mandatory 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 <text> 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	
Mandatory Value format	free text Language-dependent: this value may require translation for International/International/National/Regional procedures	
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	
Mandatory Value format	free text	
Example	<INSDQualifier_value>MCF7</INSDQualifier_value>	

6.5 Qualifier	cell_type
Definition	cell type from which the sequence was obtained
Mandatory Value format	free text Language-dependent: this value may require translation for International/International/National/Regional procedures
Example	<INSDQualifier_value>leukocyte</INSDQualifier_value>
6.6 Qualifier	chromosome
Definition	chromosome (e.g., Chromosome number) from which the sequence was obtained
Mandatory Value format	free text
Example	<INSDQualifier_value>1</INSDQualifier_value> <INSDQualifier_value>X</INSDQualifier_value>
6.7. Qualifier	circular_RNA
Definition	indicates that exons are out-of-order or overlapping because this spliced RNA product is a circular RNA (circRNA) created by backsplicing, for example when a downstream exon in the gene is located 5' of an upstream exon in the RNA product
Value format	none
Comment	should be used on features such as CDS, mRNA, tRNA and other features that are produced as a result of a backsplicing event. This qualifier should be used only when the splice event is indicated in the "join" operator, eg join(complement(69611..69724),139856..140087)
6.8 Qualifier	clone
Definition	clone from which the sequence was obtained
Mandatory Value format	free text Language-dependent: this value may require translation for International/International/National/Regional procedures
Example	<INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value>
Comment	a source feature must not contain more than one clone should qualifier; where the sequence was obtained from multiple clones it may be further described in the feature table using the feature key misc_feature and a note qualifier to specify the multiple clones.
6.9 Qualifier	clone_lib
Definition	clone library from which the sequence was obtained

Mandatory Value format free text
Language-dependent: this value may require translation for International/International/National/Regional procedures

Example <INSDQualifier_value>lambda-hIL7</INSDQualifier_value>

6.10 Qualifier codon_start

Definition indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature.

Mandatory Value format 1 or 2 or 3

Example <INSDQualifier_value>2</INSDQualifier_value>

6.11 Qualifier collected_by

Definition name of persons or institute who collected the specimen

Mandatory Value format free text
Language-dependent: this value may require translation for International/International/National/Regional procedures

Example <INSDQualifier_value>Dan Janzen</INSDQualifier_value>

6.12 Qualifier collection_date

Definition date that the specimen was collected

Mandatory Value format YYYY-MM-DD, YYYY-MM or YYYY

Example <INSDQualifier_value>1952-10-21</INSDQualifier_value>
<INSDQualifier_value>1952-10</INSDQualifier_value>
<INSDQualifier_value>1952</INSDQualifier_value>

Comment 'YYYY' is a four-digit value representing the year. 'MM' is a two-digit value representing the month. 'DD' is a two-digit value representing the day of the month.

6.13 Qualifier compare

Definition Reference details of an existing public INSD entry to which a comparison is made

Mandatory Value format [accession-number.sequence-version]

Example <INSDQualifier_value>AJ634337.1</INSDQualifier_value>

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.

6.14 Qualifier	cultivar
Definition	cultivar (cultivated variety) of plant from which sequence was obtained
Mandatory Value format	free text Language-dependent: this value may require translation for International/International/National/Regional procedures
Example	<INSDQualifier_value>Nipponbare</INSDQualifier_value> <INSDQualifier_value>Tenuifolius</INSDQualifier_value> <INSDQualifier_value>Candy Cane</INSDQualifier_value> <INSDQualifier_value>IR36</INSDQualifier_value>
Comment	'cultivar' is applied solely to products of artificial selection; use the variety qualifier for natural, named plant and fungal varieties.
6.15 Qualifier	dev_stage
Definition	if the sequence was obtained from an organism in a specific developmental stage, it is specified with this qualifier
Mandatory Value format	free text Language-dependent: this value may require translation for International/International/National/Regional procedures)
Example	<INSDQualifier_value>fourth instar larva</INSDQualifier_value>
6.16 Qualifier	direction
Definition	direction of DNA replication
Mandatory 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_value>left</INSDQualifier_value>
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.
6.17 Qualifier	EC_number
Definition	Enzyme Commission number for enzyme product of sequence
Mandatory Value format	free text
Example	<INSDQualifier_value>1.1.2.4</INSDQualifier_value> <INSDQualifier_value>1.1.2.-</INSDQualifier_value> <INSDQualifier_value>1.1.2.n</INSDQualifier_value> <INSDQualifier_value>1.1.2.n1</INSDQualifier_value>

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 may be replaced by dash "-" to indicate uncertain assignment. Symbols including an "n", e.g., "n", "n1" and so on, may 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.

6.18 Qualifier ecotype

Definition a population within a given species displaying genetically based, phenotypic traits that reflect adaptation to a local habitat

Mandatory Value Format free text
Language-dependent: this value may require translation for International/International/National/Regional procedures

Example <INSDQualifier_value>Columbia</INSDQualifier_value>

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.

6.19 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; a source feature including the environmental_sample qualifier must not include the strain qualifier.

6.20 Qualifier	exception
Definition	indicates that the coding region cannot be translated using standard biological rules
Mandatory Value format	One of the following controlled vocabulary phrases: RNA editing rearrangement required for product annotated by transcript or proteomic data
Example	<INSDQualifier_value>RNA editing</INSDQualifier_value> <INSDQualifier_value>rearrangement required for product</INSDQualifier_value>
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 corresponding conceptual translation; must not be used where transl_except qualifier would be adequate, e.g., in case of stop codon completion use.
6.21 Qualifier	frequency
Definition	frequency of the occurrence of a feature
Mandatory Value format	free text representing the proportion of a population carrying the feature expressed as a fraction Language-dependent: this value may require translation for International/International/National/Regional procedures
Example	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.22 Qualifier	function
Definition	function attributed to a sequence
Mandatory Value format	free text Language-dependent: this value may require translation for International/International/National/Regional procedures
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.
6.23 Qualifier	gene
Definition	symbol of the gene corresponding to a sequence region
Mandatory Value format	free text
Example	<INSDQualifier_value>ilvE</INSDQualifier_value>

Comment Use gene qualifier to provide the gene symbol; use standard_name qualifier to provide the full gene name.

6.24 Qualifier gene_synonym

Definition synonymous, replaced, obsolete or former gene symbol

Mandatory Value format free text
Language-dependent: this value may require translation for International/International/National/Regional procedures

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 the gene_synonym qualifier is used, a primary gene symbol must always be indicated in a gene qualifier

6.25 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 must not be used to indicate that the source of the sequence is a gamete or germ cell; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must 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)

6.26 Qualifier haplogroup

Definition name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups.

Mandatory Value format free text
Language-dependent: this value may require translation for International/International/National/Regional procedures

Example <INSDQualifier_value>H*</INSDQualifier_value>

6.27 Qualifier haplotype

Definition 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.

Mandatory Value format free text

Example <INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>

6.28 Qualifier host

Definition natural (as opposed to laboratory) host to the organism from which sequenced molecule was obtained

Mandatory Value format free text
Language-dependent: this value may require translation for International/National/Regional procedures

Example <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
<INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value>
<INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>

6.29 Qualifier identified_by

Definition name of the expert who identified the specimen taxonomically

Mandatory Value format free text
Language-dependent: this value may require translation for International/National/Regional procedures

Example <INSDQualifier_value>John Burns</INSDQualifier_value>

6.30 Qualifier isolate

Definition individual isolate from which the sequence was obtained

Mandatory Value format free text
Language-dependent: this value may require translation for International/National/Regional procedures

Example <INSDQualifier_value>Patient #152</INSDQualifier_value>
<INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>

6.31 Qualifier isolation_source

Definition describes the physical, environmental and/or local geographical
source of the biological sample from which the sequence was derived

Mandatory Value format free text
Language-dependent: this value may require translation for International/National/Regional procedures

Examples <INSDQualifier_value>rumen isolates from standard Pelleted ration-fed steer #67</INSDQualifier_value>
<INSDQualifier_value>permanent Antarctic sea

ice</INSDQualifier_value>
<INSDQualifier_value>denitrifying activated sludge from

carbon_limited continuous reactor</INSDQualifier_value>

Comment used only with the source feature key; source feature keys containing an environmental_sample qualifier should also contain an isolation_source qualifier

6.32 Qualifier

lab_host

Definition

scientific name of the laboratory host used to propagate the source

organism from which the sequenced molecule was obtained

Mandatory Value format

free text
Language-dependent: this value may require translation for International/National/Regional procedures

Example

<INSDQualifier_value>Gallus gallus</INSDQualifier_value>
<INSDQualifier_value>Gallus 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

6.33 Qualifier

lat_lon

Definition

geographical coordinates of the location where the specimen was collected

Mandatory Value format

free text - degrees latitude and longitude in format "d[d.ddd] N|S
d[dd.ddd] W|E"

Example

<INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value>
<INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>

6.34 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

6.35 Qualifier

map

Definition

genomic map position of feature

Mandatory Value format

free text

Example <INSDQualifier_value>8q12- q13</INSDQualifier_value>

6.36 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

Mandatory Value format free text
Language-dependent: this value may require translation for International/National/Regional procedures

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.

6.37 Qualifier mobile_element_type

Definition type and name or identifier of the mobile element which is described by the parent feature

Mandatory Value format <mobile_element_type>[:<mobile_element_name>]
where <mobile_element_type> is one of the following:
transposon
retrotransposon
integron
insertion sequence
non-LTR retrotransposon
SINE
MITE
LINE
other

Example <INSDQualifier_value>transposon:Tnp9</INSDQualifier_value>

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>

6.38 Qualifier mod_base

Definition abbreviation for a modified nucleotide base

Mandatory Value format modified base abbreviation chosen from this Annex, Section 2

Example <INSDQualifier_value>m5c</INSDQualifier_value>
<INSDQualifier_value>OTHER</INSDQualifier_value>

Comment specific modified nucleotides not found in paragraph 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

6.39 Qualifier mol_type

Definition molecule type of sequence

Mandatory Value 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

Example <INSDQualifier_value>genomic DNA</INSDQualifier_value>
<INSDQualifier_value>other RNA</INSDQualifier_value>

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 must be described using "genomic DNA"); ribosomal RNA genes must be described using "genomic DNA"; "rRNA" must only be used if the ribosomal RNA molecule itself has been sequenced; values "other RNA" and "other DNA" must be applied to synthetic molecules, values "unassigned DNA", "unassigned RNA" must be applied where in vivo molecule is unknown.

6.40 Qualifier ncRNA_class

Definition a structured description of the classification of the non-coding RNA described by the ncRNA parent key

Mandatory value format TYPE
where TYPE is one of the following controlled vocabulary terms or phrases:
antisense_RNA
autocatalytically_spliced_intron
circRNA
ribozyme
hammerhead_ribozyme
lncRNA
RNase_P_RNA
RNase_MRP_RNA
telomerase_RNA
guide_RNA
sgRNA
asiRNA

	sCRNA scaRNA siRNA pre_mirRNA miRNA piRNA snoRNA snRNA SRP_RNA vault_RNA Y_RNA Other
Example	<pre><INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>siRNA</INSDQualifier_value> <INSDQualifier_value>sCRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value></pre>
Comment	specific ncRNA types not yet in the ncRNA_class controlled vocabulary must be annotated by entering other as the ncRNA_class qualifier value, and providing a brief explanation of novel ncRNA_class in a note qualifier
6.41 Qualifier	note
Definition	any comment or additional information
Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
Example	<pre><INSDQualifier_value>A comment about the feature</INSDQualifier_value></pre>
6.42 Qualifier	number
Definition	a number to indicate the order of genetic elements (e.g., exons or introns) in the 5' to 3' direction
Mandatory Value format	free text (with no whitespace characters)
Example	<pre><INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value></pre>
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
6.43 Qualifier	operon
Definition	name of the group of contiguous genes transcribed into a single transcript to which that feature belongs

Mandatory value format free text

Example <INSDQualifier_value>lac</INSDQualifier_value>

6.44 Qualifier organelle

Definition type of membrane-bound intracellular structure from which the sequence was obtained

Mandatory value format One of the following controlled vocabulary terms and phrases:
chromatophore
hydrogenosome
mitochondrion
nucleomorph
plastid
mitochondrion:kinetoplast
plastid:chloroplast
plastid:apicoplast
plastid:chromoplast
plastid:cyanelle
plastid:leucoplast
plastid:proplastid,

Examples <INSDQualifier_value>chromatophore</INSDQualifier_value>
<INSDQualifier_value>hydrogenosome</INSDQualifier_value>

<INSDQualifier_value>mitochondrion</INSDQualifier_value>
<INSDQualifier_value>nucleomorph</INSDQualifier_value>
<INSDQualifier_value>plastid</INSDQualifier_value>
<INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value>
>
<INSDQualifier_value>plastid:chloroplast</INSDQualifier_value>
<INSDQualifier_value>plastid:apicoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:chromoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:cyanelle</INSDQualifier_value>
<INSDQualifier_value>plastid:leucoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:proplastid</INSDQualifier_value>

6.45 Qualifier organism

Definition 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

Mandatory value format free text
Language-dependent: this value may require translation for International/National/Regional procedures

Example <INSDQualifier_value>Homo sapiens</INSDQualifier_value>

6.46 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
Mandatory 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	<pre><INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgattttttgtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value> <INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<lt;i>gt;gtgtatcttact</INSDQualifier_value> <INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgattttttgtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value></pre>
Comment	fwd_seq and rev_seq are both mandatory; fwd_name and rev_name are both optional. Both sequences must be presented in 5' to 3' order. The sequences must be given in the symbols from Annex 1, paragraph 1, except for the modified bases, which must be enclosed within angle brackets < >. In XML, the angle brackets < and > must be substituted with < and > since they are reserved characters in XML.

6.47 Qualifier	phenotype
Definition	phenotype conferred by the feature, where phenotype is defined as a physical, biochemical or behavioural characteristic or set of characteristics
Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
Example	<INSDQualifier_value>erythromycin resistance</INSDQualifier_value>
6.48 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
Mandatory Value format	free text
Example	<INSDQualifier_value>pc589</INSDQualifier_value>
6.49 Qualifier	pop_variant
Definition	name of subpopulation or phenotype of the sample from which the sequence was derived

Mandatory Value format free text
Language-dependent: this value may require translation for

International/National/Regional procedures

Example
<INSDQualifier_value>pop1</INSDQualifier_value>
<INSDQualifier_value>Bear Paw</INSDQualifier_value>

6.50 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.

Mandatory Value format free text
Language-dependent: this value may require translation for
International/National/Regional procedures

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)

6.51 Qualifier protein_id

Definition protein sequence identification number, an integer used in a sequence listing to designate the protein sequence encoded by the coding sequence identified in the corresponding CDS feature key and translation qualifier

Mandatory Value format an integer greater than zero

Example
<INSDQualifier_value>89</INSDQualifier_value>

6.52 Qualifier proviral

Definition this qualifier is used to flag sequence obtained from a virus or phage that is integrated into the genome of another organism

Value format none

6.53 Qualifier pseudo

Definition indicates that this feature is a non-functional version of the element named by the feature key

Value format none

Comment
The qualifier pseudo should be used to describe non-functional genes that are not formally described as pseudogenes, e.g., CDS has

no translation due to other reasons than pseudogenization events. Other reasons may include sequencing or assembly errors. In order to annotate pseudogenes the qualifier pseudogene must be used, indicating the TYPE of pseudogene.

6.54 Qualifier	pseudogene
Definition	indicates that this feature is a pseudogene of the element named by the feature key
Mandatory Value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: processed unprocessed unitary allelic unknown
Example	<pre><INSDQualifier_value>processed</INSDQualifier_value> <INSDQualifier_value>unprocessed</INSDQualifier_value> <INSDQualifier_value>unitary</INSDQualifier_value> <INSDQualifier_value>allelic</INSDQualifier_value> <INSDQualifier_value>unknown</INSDQualifier_value></pre>
Comment	Definitions of TYPE values: processed - the pseudogene has arisen by reverse transcription of a mRNA into cDNA, followed by reintegration into the genome. Therefore, it has lost any intron/exon structure, and it might have a pseudo-polyA-tail. unprocessed - the pseudogene has arisen from a copy of the parent gene by duplication followed by accumulation of random mutations. The changes, compared to their functional homolog, include insertions, deletions, premature stop codons, frameshifts and a higher proportion of non-synonymous versus synonymous substitutions. unitary - the pseudogene has no parent. It is the original gene, which is functional in some species but disrupted in some way (indels, mutation, recombination) in another species or strain. allelic - a (unitary) pseudogene that is stable in the population but importantly it has a functional alternative allele also in the population. i.e., one strain may have the gene, another strain may have the pseudogene. MHC haplotypes have allelic pseudogenes. unknown - the submitter does not know the method of pseudogenization.

6.55 Qualifier	rearranged
Definition	the sequence presented in the entry has undergone somatic rearrangement as part of an adaptive immune response; it is not the unrearranged sequence that was inherited from the parental germline
Value format	none
Comment	The rearranged qualifier must not be used to annotate chromosome

rearrangements that are not involved in an adaptive immune response; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must 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)

6.56. Qualifier	recombination_class
Definition	a structured description of the classification of recombination hotspot region within a sequence
Mandatory Value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: meiotic mitotic non_allelic_homologous chromosome_breakpoint other
Example	<INSDQualifier_value>meiotic recombination</INSDQualifier_value> <INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value>
Comment	specific recombination classes not yet in the recombination_class controlled vocabulary must be annotated by entering "other" as the recombination_class qualifier value and providing a brief explanation of the novel recombination_class in a note qualifier
6.57. Qualifier	regulatory_class
Definition	a structured description of the classification of transcriptional, translational, replicational and chromatin structure related regulatory elements in a sequence
Mandatory Value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: attenuator CAAT_signal DNase_I_hypersensitive_site enhancer enhancer_blocking_element GC_signal imprinting_control_region insulator locus_control_region matrix_attachment_region minus_35_signal minus_10_signal polyA_signal_sequence promoter

	recoding_stimulatory_region recombination_enhancer replication_regulatory_region response_element ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region uORF other
Example	<INSDQualifier_value>promoter</INSDQualifier_value> <INSDQualifier_value>enhancer</INSDQualifier_value> <INSDQualifier_value>ribosome_binding_site</INSDQualifier_value>
Comment	specific regulatory classes not yet in the regulatory_class controlled vocabulary must be annotated by entering "other" as the regulatory_class qualifier value and providing a brief explanation of the novel regulatory_class in a note qualifier
6.58. Qualifier	replace
Definition	indicates that the sequence identified in a feature's location is replaced by the sequence shown in the qualifier's value; if no sequence (i.e., no value) is contained within the qualifier, this indicates a deletion
Mandatory Value format	free text
Example	<INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value></INSDQualifier_value> - for a deletion
6.59. Qualifier	ribosomal_slippage
Definition	during protein translation, certain sequences can program ribosomes to change to an alternative reading frame by a mechanism known as ribosomal slippage
Value format	none
Comment	a join operator, e.g.: [join(486..1784,1787..4810)] must be used in the CDS feature location to indicate the location of ribosomal_slippage
6.60. Qualifier	rpt_family
Definition	type of repeated sequence; "Alu" or "Kpn", for example
Mandatory Value format	free text
Example	<INSDQualifier_value>Alu</INSDQualifier_value>

6.61	Qualifier	rpt_type
	Definition	structure and distribution of repeated sequence
	Mandatory Value format	One of the following controlled vocabulary terms or phrases: tandem direct inverted flanking nested terminal dispersed long_terminal_repeat non_ltr_retrotransposon_polymeric_tract centromeric_repeat telomeric_repeat x_element_combinatorial_repeat y_prime_element other
	Example	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
	Comment	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.
6.62	Qualifier	rpt_unit_range
	Definition	location of a repeating unit expressed as a range

Mandatory Value format	<base_range> - where <base_range> is the first and last base (separated by two dots) of a repeating unit
Example	<INSDQualifier_value>202..245</INSDQualifier_value>
Comment	used to indicate the base range of the sequence that constitutes a repeating unit within the region specified by the feature keys orit and repeat_region.
6.63 Qualifier	rpt_unit_seq
Definition	identity of a repeat sequence
Mandatory Value format	free text
Example	<INSDQualifier_value>aagggc</INSDQualifier_value> <INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value> <INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value>
Comment	used to indicate the literal sequence that constitutes a repeating unit within the region specified by the feature keys orit and repeat_region
6.64 Qualifier	satellite
Definition	identifier for a satellite DNA marker, composed of many tandem repeats (identical or related) of a short basic repeated unit
Mandatory Value format	<satellite_type>[:<class>][<identifier>] - where <satellite_type> is one of the following: satellite microsatellite minisatellite
Example	<INSDQualifier_value>satellite: SlA</INSDQualifier_value> <INSDQualifier_value>satellite: alpha</INSDQualifier_value> <INSDQualifier_value>satellite: gamma III</INSDQualifier_value> <INSDQualifier_value>microsatellite: DC130</INSDQualifier_value>
Comment	many satellites have base composition or other properties that differ from those of the rest of the genome that allows them to be identified.
6.65 Qualifier	segment
Definition	name of viral or phage segment sequenced
Mandatory Value format	free text
Example	<INSDQualifier_value>6</INSDQualifier_value>
6.66 Qualifier	serotype
Definition	serological variety of a species characterized by its antigenic

	properties
Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
Example	<INSDQualifier_value>B1</INSDQualifier_value>
Comment	used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for the prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".
6.67 Qualifier	serovar
Definition	serological variety of a species (usually a prokaryote) characterized by its antigenic properties
Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
Example	<INSDQualifier_value>O157:H7</INSDQualifier_value>
Comment	used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".
6.68 Qualifier	sex
Definition	sex of the organism from which the sequence was obtained; sex is used for eukaryotic organisms that undergo meiosis and have sexually dimorphic gametes
Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
Examples	<INSDQualifier_value>female</INSDQualifier_value> <INSDQualifier_value>male</INSDQualifier_value> <INSDQualifier_value>hermaphrodite</INSDQualifier_value> <INSDQualifier_value>unisexual</INSDQualifier_value> <INSDQualifier_value>bisexual</INSDQualifier_value> <INSDQualifier_value>asexual</INSDQualifier_value> <INSDQualifier_value>monoecious</INSDQualifier_value> [or monocious] <INSDQualifier_value>dioecious</INSDQualifier_value> [or diecious]
Comment	The sex qualifier should be used (instead of mating_type qualifier) in the Metazoa, Embryophyta, Rhodophyta & Phaeophyceae; mating_type qualifier should be used (instead of sex qualifier) in the Bacteria, Archaea & Fungi; neither sex nor mating_type qualifiers should be used in the viruses; outside of the taxa listed above, mating_type qualifier should be used unless the value of the

qualifier is taken from the vocabulary given in the examples above

6.69	Qualifier	standard_name
	Definition	accepted standard name for this feature
	Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>dotted</INSDQualifier_value>
	Comment	use standard_name qualifier to give full gene name, but use gene qualifier to give gene symbol (in the above example gene qualifier value is Dt).
6.70	Qualifier	strain
	Definition	strain from which sequence was obtained
	Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>BALB/c</INSDQualifier_value>
	Comment	feature entries including a strain qualifier must not include the environmental_sample qualifier
6.71	Qualifier	sub_clone
	Definition	sub-clone from which sequence was obtained
	Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value>
	Comment	a source feature must not contain more than one sub_clone qualifier; to indicate that the sequence was obtained from multiple sub_clones, multiple sources may be further described using the feature key "misc_feature" and the qualifier "note"
6.72	Qualifier	sub_species
	Definition	name of sub-species of organism from which sequence was obtained
	Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>lactis</INSDQualifier_value>

6.73	Qualifier	sub_strain
	Definition	name or identifier of a genetically or otherwise modified strain from which sequence was obtained, derived from a parental strain (which should be annotated in the strain qualifier). sub_strain from which sequence was obtained
	Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>abis</INSDQualifier_value>
	Comment	must be accompanied by a strain qualifier in a source feature; if the parental strain is not given, the modified strain should be annotated in the strain qualifier instead of sub_strain. For example, either a strain qualifier with the value K-12 and a substrain qualifier with the value MG1655 or a strain qualifier with the value MG1655
6.74	Qualifier	tag_peptide
	Definition	base location encoding the polypeptide for proteolysis tag of tmRNA and its termination codon
	Mandatory Value format	<base_range> - where <base_range> provides the first and last base (separated by two dots) of the location for the proteolysis tag
	Example	<INSDQualifier_value>90..122</INSDQualifier_value>
	Comment	it is recommended that the amino acid sequence corresponding to the tag_peptide be annotated by describing a 5' partial CDS feature; e.g., CDS with a location of <90..122
6.75	Qualifier	tissue_lib
	Definition	tissue library from which sequence was obtained
	Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>tissue library 772</INSDQualifier_value>
6.76	Qualifier	tissue_type
	Definition	tissue type from which the sequence was obtained
	Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>liver</INSDQualifier_value>

6.77	Qualifier	transl_except
	Definition	translational exception: single codon the translation of which does not conform to genetic code defined by organism or transl_table.
	Mandatory 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 Annex 1, paragraph 3, 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'.
6.78	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
	Mandatory 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.
6.79	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 must be used only when the splice event is indicated in the "join" operator, e.g., join(complement(69611..69724),139856..140087) in the feature location
6.80 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
Mandatory Value format	contiguous string of one-letter amino acid abbreviations from this Annex paragraph 3, "X" is to be used for AA exceptions.
Example	<INSDQualifier_value>MASTFPWPYRGCASTPSLKGLIMCTW</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_table 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.
6.81 Qualifier	variety
Definition	variety (= varietas, a formal Linnaean rank) of organism from which sequence was derived.
Mandatory Value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
Example	<INSDQualifier_value>insularis</INSDQualifier_value>
Comment	use the cultivar qualifier for cultivated plant varieties, i.e., products of artificial selection; varieties other than plant and fungal variatas should be annotated via a note qualifier, e.g., with the value <INSDQualifier_value>breed:Cukorova</INSDQualifier_value>

Раздел 7 Функциональные ключи для аминокислотных последовательностей

Данная глава содержит перечень возможных функциональных ключей, которые будут использоваться в аминокислотных последовательностях. Функциональные ключи перечислены в алфавитном порядке.

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 must be annotated separately with the ACT_SITE feature key. The corresponding amino acid residue number must 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 descriptor in the feature location element is the residue numbers of the cross-linked amino acids in “x.y” format, e.g. “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 descriptor in the feature location element is the residue numbers of the linked cysteines in "x..y" format, e.g. "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
	Definition	Binding site for a metal ion.
	Mandatory qualifiers	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
	Definition	Posttranslational modification of a residue
	Mandatory qualifiers	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 full name. Examples of values for the "NOTE" qualifier: "N-acetylalanine"; "3-Hyp"; and "MeLys" or "N-6-methyllysine"
7.19	Feature Key	MOTIF
	Definition	Short (up to 20 amino acids) sequence motif of biological interest
	Optional qualifiers	note
7.20	Feature Key	MUTAGEN
	Definition	Site which has been experimentally altered by mutagenesis

	Optional qualifiers	note
7.21	Feature Key	NON_STD
	Definition	Non-standard amino acid
	Optional qualifiers	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
	Definition	The residue at an extremity of the sequence is not the terminal residue
	Optional qualifiers	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
	Definition	Extent of a nucleotide phosphate-binding region
	Mandatory qualifiers	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 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
	Definition	Secondary structure: Beta-strand; for example Hydrogen bonded beta-strand or residue in an isolated beta-bridge
	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.

7.32	Feature Key	TOPO_DOM
	Definition	Topological domain
	Optional qualifiers	note
7.33	Feature Key	TRANSMEM
	Definition	Extent of a transmembrane region
	Optional qualifiers	note
7.34	Feature Key	TRANSIT
	Definition	Extent of a transit peptide (mitochondrion, chloroplast, thylakoid, cyanelle, peroxisome etc.)
	Optional qualifiers	note
7.35	Feature Key	TURN
	Definition	Secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)
	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.
7.36	Feature Key	UNSURE
	Definition	Uncertainties in the sequence
	Optional qualifiers	note
	Comment	Used to describe region(s) of an amino acid sequence for which the authors are unsure about the sequence presentation.
7.37	Feature Key	VARIANT
	Definition	Authors report that sequence variants exist.
	Optional qualifiers	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”

Раздел 8 Квалификаторы для аминокислотных последовательностей

Данная глава содержит перечень возможных квалификаторов, которые будут использоваться в аминокислотных последовательностях.

Если формат значения представляет собой свободный текст, зависящий от языка, необходимо использовать одно из следующих значений:

- 1) элемент `INSDQualifier_value`; или же
- 2) элемент `NonEnglishQualifier_value`; или же
- 3) как элемент `INSDQualifier_value`, так и элемент `NonEnglishQualifier_value`.

Если формат значения не определяется как свободный текст, зависящий от языка, должен использоваться элемент `INSDQualifier_value`, а элемент `NonEnglishQualifier_value` не должен использоваться.

ПОЖАЛУЙСТА, ОБРАТИТЕ ВНИМАНИЕ: Любое значение квалификатора, предусмотренное для квалификатора со значением формата "free text" («свободный текст»), может потребовать перевода для международных, национальных или региональных процедур. Считается, что квалификаторы, перечисленные в следующей таблице, имеют значения свободного текста, зависящие от языка:

Таблица 6: Список квалификаторов со значениями свободного текста, зависящими от языка для аминокислотных последовательностей

Глава	Квалификатор свободного текста зависящего от языка
8.2	note
8.3	organism

8.1	Qualifier	mol_type
	Definition	In vivo molecule type of sequence
	Mandatory value format	protein
	Example	<INSDQualifier_value>protein</INSDQualifier_value>
	Comment	The "mol_type" qualifier is mandatory on the source feature key.

8.2	Qualifier	note
	Definition	Any comment or additional information
	Mandatory value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
	Example	<INSDQualifier_value>Heme (covalent)</INSDQualifier_value>

Comment	The “note” qualifier is mandatory for the feature keys: BINDING; CARBOHYD; CROSSLNK; DISULFID; DNA_BIND; DOMAIN; LIPID; METAL; MOD_RES; NP_BIND; SITE and ZN_FING
8.3 Qualifier	organism
Definition	Scientific name of the organism that provided the peptide
Mandatory value format	free text Language-dependent: this value may require translation for International/National/Regional procedures
Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
Comment	The “organism” qualifier is mandatory for the source feature key.

Раздел 9 Таблицы генетических кодов

Таблица 7 представляет Таблицы Генетических Кодов, которые будут использоваться для преобразования кодирующих последовательностей. Значением для квалификатора `trans_table` является номер, присвоенный соответствующей таблице генетического кода. Где `CDS` описывается квалификатором преобразования, но не квалификатором `transl_table`, для преобразования используется по умолчанию Standard Code (стандартный код) – 1. (Примечание: Таблицы генетических кодов 7, 8, 15 и 17-20 не существуют, поэтому эти числа отсутствуют в Таблице 7).

Table 7: Genetic Code Tables

[illegible]

[illegible]

[illegible]

31 - Blastocrithidia Nuclear Code	
AAs =	FLLSSSSYYEECCWLLLLPPPHHQQRRRIIIMTTTTNNKKSSRRVVVAAAADDEEGGGG
Starts =	-----**-----M-----
Base1 =	ttttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggggg
Base2 =	ttttccccaaaagggtttttccccaaaagggtttttccccaaaagggtttttccccaaaagggt
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
33 - Cephalodiscidae Mitochondrial UAA-Tyr Code	
AAs =	FLLSSSSYYY*CCWLLLLPPPHHQQRRRIIIMTTTTNNKKSSSKVVVVAAAADDEEGGGG
Starts =	---M-----M-----M-----M-----
Base1 =	ttttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggggg
Base2 =	ttttccccaaaagggtttttccccaaaagggtttttccccaaaagggtttttccccaaaagggt
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag

[Приложение 2 следует]