F

OMPI
ORGANISATION MONDIALE
DE LA **PROPRIÉTÉ**
INTELLECTUELLE

Comité des normes de l'OMPI (CWS)

**Septième session
Genève, 1er – 5 juillet 2019**

REVISION DE LA NORME ST.26 DE L'OMPI

Document préparé par le Bureau international

INTRODUCTION

1.   À la sixième session du Comité des normes de l'OMPI (CWS) tenue à Genève en 2018, l'équipe d'experts chargée de la norme relative aux listages des séquences (l'équipe d'experts SEQL) a proposé une série de modifications à apporter à la norme ST.26 de l'OMPI (voir le document CWS/6/16). Ces modifications portaient sur le corps du texte et les annexes I, II, III, IV et VI de la norme ST.26 de l'OMPI et comprenaient l'ajout d'une nouvelle annexe VII qui contient des tables de conversion entre la norme ST.25 de l'OMPI et la norme ST.26 de l'OMPI. Le CWS a approuvé une nouvelle version de la norme, la version 1.2, qui contenait les propositions de modifications détaillées dans le document CWS/6/16 mais aussi les modifications suivantes :

– remplacement de trois occurrences du mot "legal" par le mot "permitted";
– remplacement de 15 occurrences du mot "portion(s)" par le mot "région(s)" dans l'annexe VI intitulée "Document d'orientation"; et
– ajout d'une nouvelle phrase : La clé de caractérisation "modified_base" pour la séquence de nucléotides est également présente à la fois dans la norme ST.25 et dans la norme ST.26; cependant, le scénario 7 contient des recommandations appropriées après la première phrase dans le scénario 9 à l'annexe VII intitulée "Recommandation relative à la conversion d'un listage des séquences de la norme ST.25 à la norme ST.26".

2.   À sa sixième session, le CWS est également convenu de modifier la description de la tâche n° 44 comme suit :

–   "Fournir un appui au Bureau international en communiquant les besoins et le retour d'information des utilisateurs sur l'outil d'édition et de validation de la norme ST.26; fournir un appui au Bureau international pour les révisions à apporter en conséquence aux Instructions administratives du PCT; et préparer les révisions à apporter à la norme ST.26 de l'OMPI."

3.   En conséquence, la version la plus récente de la norme ST.26 de l'OMPI, la version 1.2, a été publiée en février 2019.

4.   L'équipe d'experts s'est réunie deux fois pour examiner les révisions à apporter à la norme ST.26 de l'OMPI : lors d'une réunion physique pendant la sixième session du CWS en octobre 2018 et d'une conférence en ligne en avril 2019.

SYNTHESE DE LA PROPOSITION DE REVISION

Modifications d'ordre rédactionnel

5.   Dans le cadre de la tâche n° 44, l'équipe d'experts SEQL a examiné la version de la norme ST.26 de l'OMPI actuellement publiée, la version 1.2. À cette occasion, elle a identifié une série de modifications d'ordre rédactionnel à apporter, comme la suppression d'espaces inutiles ou d'erreurs typographiques. D'autres corrections étaient également nécessaires pour s'assurer que ce document respectait le guide de style de l'OMPI, par exemple que l'abréviation latine "e.g." était utilisée systématiquement pour remplacer "for example".

6.   Ces corrections sont surlignées en vert dans l'annexe au présent document (annexe I de la norme ST.26 de l'OMPI) et seront présentées au CWS pour examen pendant la septième session. L'équipe d'experts a identifié des modifications d'ordre rédactionnel dans les annexes I à VII de la norme ST.26 de l'OMPI, même si seule l'annexe I accompagne le présent document.

Changements de fond

7.   L'équipe d'experts SEQL propose de réviser les annexes I et VII de la norme ST.26 de l'OMPI en leur apportant les modifications suivantes; dans ces annexes, les ajouts sont surlignés en jaune, et les suppressions en violet. Aucune modification de fond du corps du texte de la norme ST.26 de l'OMPI n'est proposée :

a)   Mise à jour de l'annexe I, tableau 9, pour insérer les mises à jour contenues dans la version 10.8 du tableau des caractéristiques INSDC;

b)   Dans l'annexe I, section 5.27, ajouter les qualificateurs facultatifs supplémentaires suivants :

–   function
–   gene
–   gene_synonym
–   map

c)   Dans l'annexe I, section 5.33, ajouter les qualificateurs facultatifs supplémentaires suivants :

–   allele
–   direction
–   gene
–   gene_synonym
–   map
–   note
–   standard_name

d)   Dans l'annexe I, section 5.43, ajouter le qualificateur facultatif supplémentaire suivant :

–   operon

e)   Dans l'annexe I, section 6.16, ajouter les lignes supplémentaires suivantes :

–   dans la colonne "Example" :
<INSDQualifier_value>1.1.2.n1</INSDQualifier_value>; et
–   dans la colonne "Comment", ajouter le texte suivant : "Symbols including an "n", e.g. "n", "n1" and so on."

f)   Mise à jour de l'annexe VII, troisième paragraphe, "Recommandations relatives à l'ajout ou la suppression éventuels d'éléments", première phrase : remplacer "conversion" par "transformation". Le terme "conversion" implique qu'il existe une correspondance unique entre les composants, ce qui est techniquement incorrect.

g)   Mise à jour de l'annexe VII, scénario 8, numéro 24 : remplacer le terme "SITE" par "REGION". Ce changement a été proposé par les développeurs qui ont détecté un conflit entre la spécification fonctionnelle décrivant l'importation de séquences ST.25 et cet exemple; Instances XML dans des fichiers distincts.

8.   L'équipe d'experts SEQL propose également, pour s'aligner sur les propositions de modifications à apporter à d'autres normes de l'OMPI, que le contenu de l'annexe III de la norme ST.26 de l'OMPI et l'appendice à l'annexe VI de la norme ST.26, qui sont des instances XML, soient fournis dans deux fichiers distincts et que la norme contienne plutôt un lien vers ces fichiers. Cela devrait rendre ces exemples plus accessibles aux lecteurs de la norme.

9.   Pour de plus amples informations sur ces changements, veuillez vous reporter à l'annexe.

10.   Le CWS est invité

a)   à prendre note du contenu du présent document,

b)   à examiner et éventuellement approuver la proposition de révision de la norme ST.26 de l'OMPI décrite aux paragraphes 4 à 6 et reproduite dans les annexes I et II du présent document et

c) à examiner et éventuellement approuver la suggestion de fournir le contenu de l'annexe III et l'appendice à l'annexe VI de la norme ST.26 de l'OMPI dans deux fichiers distincts, liés à la norme, comme décrit au paragraphe 7.

[L'annexe (annexe I de la norme ST.26) suit]

NORME ST.26 – ANNEXE I

VOCABULAIRE CONTROLE

Version 1.23

Projet final

Adoptée par le Comité des normes de l'OMPI (CWS) à sa cinquième session le 2 juin 2017.Proposition présentée par l'équipe d'experts SEQL pour examen et adoption par le CWS à sa septième session

TABLE DES MATIERES

SECTION 1 : LISTE DES NUCLÉOTIDES

Les codes des bases nucléotidiques à utiliser dans les listages des séquences sont présentés dans le tableau 1. Lorsqu'il n'est pas accompagné d'une description supplémentaire, le symbole "t" désigne la thymine dans de l'ADN et l'uracile dans de l'ARN. Lorsqu'il convient d'employer un symbole ambigu (représentant deux bases nucléotidiques possibles ou plus), il faut choisir le symbole le plus restrictif. Si, par exemple, une base nucléotidique dans une position quelconque pouvait être désignée par "a ou g", il faut employer "r" au lieu de "n". Le symbole "n" sera considéré comme "a ou c ou g ou t/u" s'il n'est pas accompagné d'une description supplémentaire.

Tableau 1 : Liste des nucléotides

Symbole	Nucléotide
a	adenine
c	cytosine
g	guanine
t	thymine dans de l'ADN/uracile dans de l'ARN (t/u)
m	a ou c
r	a ou g
w	a ou t/u
s	c ou g
y	c ou t/u
k	g ou t/u
v	a ou c ou g; et non t/u
h	a ou c ou t/u; et non g
d	a ou g ou t/u; et non c
b	c ou g ou t/u; et non a
n	a ou c ou g ou t/u; "unknown" ou "other"

SECTION 2 : LISTE DES NUCLÉOTIDES MODIFIÉS

Les abréviations indiquées dans le tableau 2 sont les seules valeurs autorisées pour le qualificateur mod_base. Lorsqu'un nucléotide modifié particulier ne figure pas dans le tableau ci-après, il doit prendre pour valeur l'abréviation "OTHER". Si l'abréviation est "OTHER", le nom complet non abrégé de la base modifiée doit être indiqué dans un qualificateur du type "note". Les abréviations indiquées dans le tableau 2 ne doivent pas être employées dans la séquence elle-même.

Tableau 2 : Liste des nucléotides modifiés

Abréviation	Nucléotide modifié
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxylmethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
dhu	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta-D-galactosylqueuosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m4c	N4-methylcytosine
m5c	5-methylcytidine

Abréviation	Nucléotide modifié
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methylaminomethyl-2-thiouridine
man q	beta-D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
mv	uridine-5-oxoacetic acid-methylester
o5u	uridine-5-oxyacetic acid (v)
osyw	wybutoxosine
p	pseudouridine
q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
m5u	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u
OTHER	(nécessite un qualificateur "note")

SECTION 3 : LISTE DES ACIDES AMINÉS

Les codes des acides aminés à employer dans la séquence sont présentés dans le tableau 3. Lorsqu'il convient d'employer un symbole ambigu (représentant deux acides aminés possibles ou plus), il faut choisir le symbole le plus restrictif. Si, par exemple, un acide aminé à une position quelconque pouvait être un acide aspartique ou une asparagine, il faut employer le symbole "B" au lieu de "X". Le symbole "X" sera considéré comme l'équivalent de l'un des symboles "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" ou "V" s'il n'est pas accompagné d'une description supplémentaire.

Tableau 3 : Liste des acides aminés

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

SECTION 4 : LISTE DES ACIDES AMINÉS MODIFIÉS ~~ET INHABITUELS~~

Le tableau 4 indique les seules abréviations autorisées pour un acide aminé modifié ~~ou inhabituel~~ dans le qualificateur obligatoire du type "NOTE" pour les clés de caractérisation "MOD_RES" ou "SITE". La valeur du qualificateur du type "NOTE" doit être soit une abréviation indiquée dans ce tableau, s'il y a lieu, soit le nom complet non abrégé de l'acide aminé modifié. Les abréviations (ou les noms complets) indiquées dans ce tableau ne doivent pas être employées dans la séquence elle-même.

Tableau 4 : Liste des acides aminés modifiés ~~et inhabituels~~

Abréviation	Acide aminé modifié ~~ou inhabituel~~
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminoproprionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2′-Diaminopimelic acid
Dpr	2,3-Diaminoproprionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
aIle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
MeIle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

SECTION 5 : CLÉS DE CARACTÉRISATION POUR LES SÉQUENCES ~~NUCLÉIQUES~~ DE NUCLÉOTIDES

La présente section donne la liste des clés de caractérisation qui peuvent être employées pour les séquences de nucléotides, ainsi qu'une liste des qualificateurs obligatoires et facultatifs. Les clés de caractérisation sont présentées dans l'ordre alphabétique. Sauf indication contraire, elles peuvent être employées soit pour l'ADN, soit pour l'ARN sous "Molecule scope". Certaines Feature Keys peuvent être utilisées avec des séquences artificielles pour compléter le "organism scope" indiqué.

Les noms des clés de caractérisation doivent être employés dans l'instance XML du listage des séquences exactement comme ils apparaissent à la suite de "Feature key" dans les descriptions ci-après, à l'exception des clés de caractérisation 3'UTR et 5'UTR. On se reportera dans la description à "Comment" correspondant aux clés de caractérisation 3'UTR et 5'UTR.

5.1.	Feature Key	C_region
	Definition	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain
	Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes

5.2.	Feature Key	CDS
	Definition	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature may include amino acid conceptual translation
	Optional qualifiers	allele codon_start EC_number exception function gene gene_synonym map note number operon product protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_table trans_splicing
	Comment	codon_start qualifier has valid value of 1 or 2 or 3, indicating the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature; transl_table defines the genetic code table used if other than the Standard or universal genetic code table; genetic code exceptions outside the range of the specified tables are reported in transl_except qualifier; only one of the qualifiers translation, pseugogene 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

          Organism scope          eukaryotes
```

```
5.11. Feature Key              mat_peptide

          Definition           mature peptide or protein coding sequence; coding sequence for the mature or final
                               peptide or protein product following post-translational modification; the location
                               does not include the stop codon (unlike the corresponding CDS)

          Optional qualifiers  allele
                               EC_number
                               function
                               gene
                               gene_synonym
                               map
                               note
                               product
                               pseudo
                               pseudogene
                               standard_name
```

```
5.12. Feature Key              misc_binding

          Definition           site in nucleic acid which covalently or non-covalently binds another moiety that
                               cannot be described by any other binding key (primer_bind or protein_bind)

          Mandatory qualifiers bound_moiety

          Optional qualifiers  allele
                               function
                               gene
                               gene_synonym
                               map
                               note

          Comment              note that the regulatory feature key and regulatory_class qualifier with the
                               value"ribosome_binding_site" 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
                                  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
                                  pseudogene
                                  standard_name

    ==Organism scope==   ==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
```

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
                                 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>.

SECTION 6 : QUALIFICATEURS POUR LES SÉQUENCES ~~NUCLÉIQUES~~ DE NUCLÉOTIDES

La présente section donne la liste des qualificateurs à employer pour les caractéristiques dans les séquences de nucléotides. Les qualificateurs sont présentés dans l'ordre alphabétique.

Lorsqu'un Value format de "none" est indiqué dans la description d'un qualifier (par exemple germline), il ne faut pas utiliser l'élément `INSDQualifier_value`.

N.B. : Toute qualifier value indiquée pour un qualificateur avec un format de valeur "free text" peut devoir être traduite aux fins des procédures nationales ou régionales.

6.1.	Qualifier	`allele`
	Definition	name of the allele for the given gene
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	`<INSDQualifier_value>adh1-1</INSDQualifier_value>`
	Comment	all gene-related features (exon, CDS etc) for a given gene should share the same allele qualifier value; the allele qualifier value must, by definition, be different from the gene qualifier value; when used with the variation feature key, the allele qualifier value should be that of the variant.

6.2.	Qualifier	`anticodon`
	Definition	location of the anticodon of tRNA and the amino acid for which it codes
	Value format	(pos:<location>,aa:<amino_acid>,seq:<text>) where <location> is the position of the anticodon and <amino_acid> is the three letter abbreviation for the amino acid encoded and <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
	Value format	free text (NOTE: this value may require translation for 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
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	`<INSDQualifier_value>MCF7</INSDQualifier_value>`

6.5.	Qualifier	cell_type
	Definition	cell type from which the sequence was obtained
	Value format	free text (NOTE: this value may require translation for 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
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>1</INSDQualifier_value> <INSDQualifier_value>X</INSDQualifier_value>

6.7.	Qualifier	clone
	Definition	clone from which the sequence was obtained
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value>
	Comment	a source feature must not contain more than one clone 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.

6.8.	Qualifier	clone_lib
	Definition	clone library from which the sequence was obtained
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>lambda-hIL7</INSDQualifier_value>

6.9.	Qualifier	codon_start
	Definition	indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature.
	Value format	1 or 2 or 3
	Example	<INSDQualifier_value>2</INSDQualifier_value>

6.10.	Qualifier	collected_by
	Definition	name of persons or institute who collected the specimen
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>Dan Janzen</INSDQualifier_value>

6.11.	Qualifier	collection_date
	Definition	date that the specimen was collected.
	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.

6.12.	Qualifier	compare
	Definition	Reference details of an existing public INSD entry to which a comparison is made
	Value format	[accession-number.sequence-version]
	Example	\<INSDQualifier_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.

6.13.	Qualifier	cultivar
	Definition	cultivar (cultivated variety) of plant from which sequence was obtained
	Value format	free text (NOTE: this value may require translation for 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.

6.14.	Qualifier	dev_stage
	Definition	if the sequence was obtained from an organism in a specific developmental stage, it is specified with this qualifier
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	\<INSDQualifier_value>fourth instar larva\</INSDQualifier_value>

6.15.	Qualifier	direction
	Definition	direction of DNA replication
	Value format	left, right, or both where left indicates toward the 5' end of the sequence (as presented) and right indicates toward the 3' end
	Example	<INSDQualifier_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.

6.16.	Qualifier	EC_number
	Definition	Enzyme Commission number for enzyme product of sequence
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	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 symbol "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.

6.17.	Qualifier	ecotype
	Definition	a population within a given species displaying genetically based, phenotypic traits that reflect adaptation to a local habitat
	Value Format	free text (NOTE: this value may require translation for 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.

6.18.	Qualifier	environmental_sample
	Definition	identifies sequences derived by direct molecular isolation from a bulk environmental DNA sample (by PCR with or without subsequent cloning of the product, DGGE, or other anonymous methods) with no reliable identification of the source organism. Environmental samples include clinical samples, gut contents, and other sequences from anonymous organisms that may be associated with a particular host. They do not include endosymbionts that can be reliably recovered from a particular host, organisms from a readily identifiable but uncultured field sample (e.g., many cyanobacteria), or phytoplasmas that can be reliably recovered from diseased plants (even though these cannot be grown in axenic culture)
	Value format	none
	Comment	used only with the source feature key; source feature keys containing the environmental_sample qualifier should also contain the isolation_source qualifier; a source feature including the environmental_sample qualifier must not include the strain qualifier.

6.19.	Qualifier	exception
	Definition	indicates that the coding region cannot be translated using standard biological rules
	Value format	One of the following controlled vocabulary phrases: RNA editing rearrangement required for product annotated by transcript or proteomic data
	Example	\<INSDQualifier_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.

6.20.	Qualifier	frequency
	Definition	frequency of the occurrence of a feature
	Value format	free text representing the proportion of a population carrying the feature expressed as a fraction (NOTE: this value may require translation for National/Regional procedures)
	Example	\<INSDQualifier_value>23/108\</INSDQualifier_value> \<INSDQualifier_value>1 in 12\</INSDQualifier_value> \<INSDQualifier_value>0.85\</INSDQualifier_value>

6.21.	Qualifier	function
	Definition	function attributed to a sequence
	Value format	free text (NOTE: this value may require translation for 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.

6.22.	Qualifier	gene
	Definition	symbol of the gene corresponding to a sequence region
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	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.

6.23.	Qualifier	gene_synonym
	Definition	synonymous, replaced, obsolete or former gene symbol
	Value format	free text (NOTE: this value may require translation for 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

6.24.	Qualifier	germline
	Definition	the sequence presented has not undergone somatic rearrangement as part of an adaptive immune response; it is the unrearranged sequence that was inherited from the parental germline
	Value format	none
	Comment	germline qualifier 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)

6.25.	Qualifier	haplogroup
	Definition	name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups.
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>H*</INSDQualifier_value>

6.26.	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.
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>

6.27.	Qualifier	host
	Definition	natural (as opposed to laboratory) host to the organism from which sequenced molecule was obtained
	Value format	free text (NOTE: this value may require translation for 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>

6.28.	Qualifier	identified_by
	Definition	name of the expert who identified the specimen taxonomically
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>John Burns</INSDQualifier_value>

6.29.	Qualifier	isolate
	Definition	individual isolate from which the sequence was obtained
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>Patient #152</INSDQualifier_value> <INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>

6.30.	Qualifier	isolation_source
	Definition	describes the physical, environmental and/or local geographical source of the biological sample from which the sequence was derived
	Value format	free text (NOTE: this value may require translation for 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

6.31.	Qualifier	lab_host
	Definition	scientific name of the laboratory host used to propagate the source organism from which the sequenced molecule was obtained
	Value format	free text (NOTE: this value may require translation for 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

6.32.	Qualifier	lat_lon
	Definition	geographical coordinates of the location where the specimen was collected
	Value format	free text – degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E" (NOTE: this value may require translation for National/Regional procedures)
	Example	\<INSDQualifier_value\>47.94 N 28.12 W\</INSDQualifier_value\> \<INSDQualifier_value\>45.0123 S 4.1234 E\</INSDQualifier_value\>

6.33.	Qualifier	macronuclear
	Definition	if the sequence shown is DNA and from an organism which undergoes chromosomal differentiation between macronuclear and micronuclear stages, this qualifier is used to denote that the sequence is from macronuclear DNA
	Value format	none

6.34.	Qualifier	map
	Definition	genomic map position of feature
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	\<INSDQualifier_value\>8q12-q13\</INSDQualifier_value\>

6.35.	Qualifier	mating_type
	Definition	mating type of the organism from which the sequence was obtained; mating type is used for prokaryotes, and for eukaryotes that undergo meiosis without sexually dimorphic gametes
	Value format	free text (NOTE: this value may require translation for 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.

6.36.	Qualifier	mobile_element_type
	Definition	type and name or identifier of the mobile element which is described by the parent feature
	Value format	<mobile_element_type>[:<mobile_element_name>] where <mobile_element_type> is one of the following: transposon retrotransposon integron insertion sequence non-LTR retrotransposon SINE MITE LINE other
	Example	<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>

6.37.	Qualifier	mod_base
	Definition	abbreviation for a modified nucleotide base
	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 Section 2 of this Annex are annotated by entering OTHER as the value for the mod_base qualifier and including a note qualifier with the full name of the modified base as its value

6.38.	Qualifier	mol_type
	Definition	molecule type of sequence
	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.

6.39.	Qualifier	ncRNA_class
	Definition	a structured description of the classification of the non-coding RNA described by the ncRNA parent key
	Value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: antisense_RNA autocatalytically_spliced_intron ribozyme hammerhead_ribozyme lncRNA RNase_P_RNA RNase_MRP_RNA telomerase_RNA guide_RNA sgRNA rasiRNA scRNA scaRNA siRNA pre_miRNA miRNA piRNA snoRNA snRNA SRP_RNA vault_RNA Y_RNA other
	Example	<INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>siRNA</INSDQualifier_value> <INSDQualifier_value>scRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value>
	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

6.40.	Qualifier	note
	Definition	any comment or additional information
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>A comment about the feature</INSDQualifier_value>

6.41.	Qualifier	number
	Definition	a number to indicate the order of genetic elements (e.g. exons or introns) in the 5' to 3' direction
	Value format	free text (with no whitespace characters) (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value>
	Comment	text limited to integers, letters or combination of integers and/or letters represented as a data value that contains no whitespace characters; any additional terms should be included in a standard_name qualifier. Example: a number qualifier with a value of 2A and a standard_name qualifier with a value of "long"

6.42.	Qualifier	operon
	Definition	name of the group of contiguous genes transcribed into a single transcript to which that feature belongs
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>lac</INSDQualifier_value>

6.43.	Qualifier	organelle
	Definition	type of membrane-bound intracellular structure from which the sequence was obtained
	Value format	One of the following controlled vocabulary terms and phrases: chromatophore hyrogenosome 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>

6.44.	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
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>

6.45.	Qualifier	PCR_primers
	Definition	PCR primers that were used to amplify the sequence. A single PCR_primers qualifier should contain all the primers used for a single PCR reaction. If multiple forward or reverse primers are present in a single PCR reaction, multiple sets of fwd_name/fwd_seq or rev_name/rev_seq values will be present
	Value format	[fwd_name: XXX1,]fwd_seq: xxxxx1,[fwd_name: XXX2,]fwd_seq: xxxxx2, [rev_name: YYY1,]rev_seq: yyyyy1,[rev_name: YYY2,]rev_seq: yyyyy2
	Example	<INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgattttttggtcayccwgaagt,rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value> <INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i>gtgtatcttact</INSDQualifier_value> <INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgattttttggtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value>
	Comment	fwd_seq and rev_seq are both mandatory; fwd_name and rev_name are both optional. Both sequences must be presented in 5'>3' order. The sequences must be given in the symbols from Section 1 of this Annex, 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.

6.46.	Qualifier	phenotype
	Definition	phenotype conferred by the feature, where phenotype is defined as a physical, biochemical or behavioural characteristic or set of characteristics
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>erythromycin resistance</INSDQualifier_value>

6.47.	Qualifier	plasmid
	Definition	name of naturally occurring plasmid from which the sequence was obtained, where plasmid is defined as an independently replicating genetic unit that cannot be described by chromosome or segment qualifiers
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>pC589</INSDQualifier_value>

6.48.	Qualifier	pop_variant
	Definition	name of subpopulation or phenotype of the sample from which the sequence was derived
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>

6.49.	Qualifier	product
	Definition	name of the product associated with the feature, e.g. the mRNA of an mRNA feature, the polypeptide of a CDS, the mature peptide of a mat_peptide, etc.
	Value format	free text (NOTE: this value may require translation for 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)

6.50.	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
	Value format	an integer greater than zero
	Example	<INSDQualifier_value>89</INSDQualifier_value>

6.51.	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

6.52.	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.

6.53.	Qualifier	pseudogene
	Definition	indicates that this feature is a pseudogene of the element named by the feature key
	Value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: processed unprocessed unitary allelic unknown
	Example	\<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>
	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 is 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.

6.54.	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)

6.55.	Qualifier	recombination_class
	Definition	a structured description of the classification of recombination hotspot region within a sequence
	Value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: meiotic mitotic_recombination non_allelic_homologous_recombination_region 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

6.56.	Qualifier	regulatory_class
	Definition	a structured description of the classification of transcriptional, translational, replicational and chromatin structure related regulatory elements in a sequence
	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 replication_regulatory_region response_element polyA_signal_sequence promoter ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region 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

6.57.	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
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	`<INSDQualifier_value>a</INSDQualifier_value>` `<INSDQualifier_value></INSDQualifier_value>` – for a deletion

6.58.	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

6.59.	Qualifier	rpt_family
	Definition	type of repeated sequence; "Alu" or "Kpn", for example
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	`<INSDQualifier_value>Alu</INSDQualifier_value>`

6.60.	Qualifier	rpt_type
	Definition	structure and distribution of repeated sequence

	Value format	One of the following controlled vocabulary terms or phrases: tandem direct inverted flanking nested <mark>terminal</mark> 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.

6.61.	Qualifier	rpt_unit_range
	Definition	location of a repeating unit expressed as a range
	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.

6.62.	Qualifier	rpt_unit_seq
	Definition	identity of a repeat sequence
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	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

6.63.	Qualifier	satellite
	Definition	identifier for a satellite DNA marker, compose of many tandem repeats (identical or related) of a short basic repeated unit
	Value format	\<satellite_type>[:\<class>][\<identifier>] – where \<satellite_type> is one of the following: satellite; microsatellite; minisatellite
	Example	\<INSDQualifier_value>satellite: S1a\</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.

6.64.	Qualifier	segment
	Definition	name of viral or phage segment sequenced
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	\<INSDQualifier_value>6\</INSDQualifier_value>

6.65.	Qualifier	serotype
	Definition	serological variety of a species characterized by its antigenic properties
	Value format	free text (NOTE: this value may require translation for 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".

6.66.	Qualifier	serovar
	Definition	serological variety of a species (usually a prokaryote) characterized by its antigenic properties
	Value format	free text (NOTE: this value may require translation for 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".

6.67.	Qualifier	sex
	Definition	sex of the organism from which the sequence was obtained; sex is used for eukaryotic organisms that undergo meiosis and have sexually dimorphic gametes
	Value format	free text (NOTE: this value may require translation for 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 monecious] <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

6.68.	Qualifier	standard_name
	Definition	accepted standard name for this feature
	Value format	free text (NOTE: this value may require translation for 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).

6.69.	Qualifier	strain
	Definition	strain from which sequence was obtained
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>BALB/c</INSDQualifier_value>
	Comment	feature entries including a strain qualifier must not include the environmental_sample qualifier

6.70.	Qualifier	sub_clone
	Definition	sub-clone from which sequence was obtained
	Value format	free text (NOTE: this value may require translation for 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"

6.71.	Qualifier	sub_species
	Definition	name of sub-species of organism from which sequence was obtained
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>lactis</INSDQualifier_value>

6.72.	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
	Value format	free text (NOTE: this value may require translation for 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

6.73.	Qualifier	tag_peptide
	Definition	base location encoding the polypeptide for proteolysis tag of tmRNA and its termination codon
	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

6.74.	Qualifier	tissue_lib
	Definition	tissue library from which sequence was obtained
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>tissue library 772</INSDQualifier_value>

6.75.	Qualifier	tissue_type
	Definition	tissue type from which the sequence was obtained
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>liver</INSDQualifier_value>

6.76.	Qualifier	transl_except
	Definition	translational exception: single codon the translation of which does not conform to genetic code defined by organism or transl_table.
	Value format	(pos:location, aa:<amino_acid>) where <amino_acid> is the three letter abbreviation for the amino acid coded by the codon at the base_range position
	Example	<INSDQualifier_value>(pos:213..215, aa:Trp) </INSDQualifier_value> <INSDQualifier_value>(pos:462..464, aa:OTHER) </INSDQualifier_value> <INSDQualifier_value>(pos:1017, aa:TERM) </INSDQualifier_value> <INSDQualifier_value>(pos:2000..2001, aa:TERM) </INSDQualifier_value>
	Comment	if the amino acid is not one of the specific amino acids listed in Section 3 of this Annex, use OTHER as <amino_acid> and provide the name of the unusual amino acid in a note qualifier; for modified amino-acid selenocysteine use three letter abbreviation 'Sec' (one letter symbol 'U' in amino-acid sequence) for <amino _acid>; for modified amino-acid pyrrolysine use three letter abbreviation 'Pyl' (one letter symbol 'O' in amino-acid sequence) for <amino _acid>; for partial termination codons where TAA stop codon is completed by the addition of 3' A residues to the mRNA either a single base_position or a base_range is used for the location, see the third and fourth examples above, in conjunction with a note qualifier indicating 'stop codon completed by the addition of 3' A residues to the mRNA'.

6.77.	Qualifier	transl_table
	Definition	definition of genetic code table used if other than universal or standard genetic code table. Tables used are described in this Annex
	Value format	<integer> where <integer> is the number assigned to the genetic code table
	Example	<INSDQualifier_value>3</INSDQualifier_value> – example where the yeast mitochondrial code is to be used
	Comment	if the transl_table qualifier is not used to further annotate a CDS feature key, then the CDS is translated using the Standard Code (i.e. Universal Genetic Code). Genetic code exceptions outside the range of specified tables are reported in transl_except qualifiers.

6.78.	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

6.79.	Qualifier	translation
	Definition	one-letter abbreviated amino acid sequence derived from either the standard (or universal) genetic code or the table as specified in a transl_table qualifier and as determined by an exception in the transl_except qualifier
	Value format	contiguous string of one-letter amino acid abbreviations from Section 3 of this Annex, "X" is to be used for AA exceptions.
	Example	<INSDQualifier_value>MASTFPPWYRGCASTPSLKGLIMCTW</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.

6.80.	Qualifier	variety
	Definition	variety (= varietas, a formal Linnaean rank) of organism from which sequence was derived.
	Value format	free text (NOTE: this value may require translation for 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>

SECTION 7 : CLÉS DE CARACTÉRISATION POUR LES SÉQUENCES D'ACIDES AMINÉS

La présente section contient la liste des clés de caractérisation pouvant être employées pour les séquences d'acides aminés. Les clés de caractérisation sont présentées dans l'ordre alphabétique.

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 descriptors in the feature location element are the residue numbers of the cross-linked amino acids in conjunction with the "join" location operator, e.g. "join(42,50)." The NOTE qualifier indicates the nature of the cross-link; at least specifying the name of the conjugate and the identity of the two amino acids involved. Examples of values for the "NOTE" qualifier: "Isoglutamyl cysteine thioester (Cys-Gln);" "Beta-methyllanthionine (Cys-Thr);" and "Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin)"

7.10.	Feature Key	DISULFID
	Definition	Disulfide bond
	~~Optional~~ Mandatory	NOTE
	Comment	For an interchain disulfide bond, the location descriptor in the feature location element is the residue number of the cysteine linked to the other protein. For an intrachain cross-link, the location descriptors in the feature location element are the residue numbers of the linked cysteines in conjunction with the "join" location operator, e.g. "join(42,50)". For interchain disulfide bonds, the NOTE qualifier indicates the nature of the cross-link, by identifying the other protein, for example, "Interchain (between A and B chains)"

7.11.	Feature Key	DNA_BIND
	Definition	Extent of a DNA-binding region
	Mandatory qualifiers	NOTE
	Comment	The nature of the DNA-binding region is given in the NOTE qualifier. Examples of values for the "NOTE" qualifier: "Homeobox" and "Myb 2"

7.12.	Feature Key	DOMAIN
	Definition	Extent of a domain, which is defined as a specific combination of secondary structures organized into a characteristic three-dimensional structure or fold
	Mandatory qualifiers	NOTE
	Comment	The domain type is given in the NOTE qualifier. Where several copies of a domain are present, the domains are numbered. Examples of values for the "NOTE" qualifier: "Ras-GAP" and "Cadherin 1"

7.13.	Feature Key	HELIX
	Definition	Secondary structure: Helices, for example, Alpha-helix; 3(10) helix; or Pi-helix
	Optional qualifiers	NOTE
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.

7.14.	Feature Key	INIT_MET
	Definition	Initiator methionine
	Optional qualifiers	NOTE
	Comment	The location descriptor in the feature location element is "1". This feature key indicates the N-terminal methionine is cleaved off. This feature is not used when the initiator methionine is not cleaved off.

7.15.	Feature Key	INTRAMEM
	Definition	Extent of a region located in a membrane without crossing it
	Optional qualifiers	NOTE

7.16.	Feature Key	LIPID
	Definition	Covalent binding of a lipid moiety
	Mandatory qualifiers	NOTE
	Comment	The chemical nature of the bound lipid moiety is given in the NOTE qualifier, indicating at least the name of the lipidated amino acid. Examples of values for the "NOTE" qualifier: "N-myristoyl glycine"; "GPI-anchor amidated serine" and "S-diacylglycerol cysteine."

7.17.	Feature Key	METAL
	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 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 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"

SECTION 8 : QUALIFICATEURS POUR LES SÉQUENCES D'ACIDES AMINÉS

La présente section donne la liste des qualificateurs pouvant être utilisés pour les séquences d'acides aminés.

N.B. : Toute qualifier value indiquée pour un qualificateur avec un format de valeur "free text" peut devoir être traduite aux fins des procédures nationales ou régionales.

8.1.	Qualifier	MOL_TYPE
	Definition	In vivo molecule type of sequence
	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
	Value format	free text (NOTE: this value may require translation for 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 and ZN_FING

8.3.	Qualifier	ORGANISM
	Definition	Scientific name of the organism that provided the peptide
	Value format	free text (NOTE: this value may require translation for National/Regional procedures)
	Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
	Comment	The "ORGANISM" qualifier is mandatory for the SOURCE feature key.

SECTION 9 : TABLEAUX DU CODE GÉNÉTIQUE

Le tableau 5 reproduit les tableaux du code génétique à utiliser pour la traduction des séquences de codage. La valeur du qualificateur trans_table est le numéro attribué au tableau du code génétique correspondant. Lorsqu'une caractéristique CDS est décrite par un qualificateur de type traduction, mais non un qualificateur transl_table, on utilise par défaut le 1 – Code normalisé pour la traduction. (Note : les tableaux du code génétique 7, 8, 15 et 17 à 20 n'existant pas, ces numéros n'apparaissent pas dans le tableau 5.)

Tableau 5 : Tableaux du code génétique

1 – Code normalisé

```
  AAs  = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M---------------M----------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

2 – Code mitochondrial des vertébrés

```
  AAs  = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSS**VVVVAAAADDEEGGGG
Starts = ----------------------------MMMM--------------M-----------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

3 – Code mitochondrial des levures

```
  AAs  = FFLLSSSSYY**CCWWTTTTPPPPHHQQRRRRIIMMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = --------------------------------MM--------------M------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

4 – Code mitochondrial des moisissures, des protozoaires et des cœlentérés et code des mycoplasmes/spiroplasmes

```
  AAs  = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = --MM---------------M-----------MMMM--------------M------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

5 – Code mitochondrial des invertébrés

```
  AAs  = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSSSSVVVVAAAADDEEGGGG
Starts = ---M----------------------MMMM--------------M------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

6 – Code génétique nucléaire des ciliés, des dasycladacées et de l'hexamita

```
  AAs  = FFLLSSSSYYQQCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = -----------------------------M----------------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

9 – Code mitochondrial des échinodermes et des platodes

```
AAs    = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts = --------------------------------M---------------M-----------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

10 – Code génétique nucléaire des euplotides

```
AAs    = FFLLSSSSYY**CCCWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = --------------------------------M---------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

11 – Code des plastides, des bactéries, <mark>**des archées**</mark> et des végétaux

```
AAs    = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M----------------M-----------MMMM---------------M-----------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

12 – Code génétique nucléaire alternatif des levures

```
AAs    = FFLLSSSSYY**CC*WLLLSPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ------------------M---------------M--------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

13 – Code mitochondrial des ascidiens

```
AAs    = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSSGGVVVVAAAADDEEGGGG
Starts = ---M----------------------------MM---------------M-----------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

14 – Code mitochondrial alternatif des platodes

```
AAs    = FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts = --------------------------------M--------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
Starts = --------------------------------M--------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

16 – Code mitochondrial des chlorophycées

```
AAs    = FFLLSSSSYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = --------------------------------M--------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

21 – Code mitochondrial des trématodes

```
  AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts = ----------------------------------M---------------M------------
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

22 – Code mitochondrial de *Scenedesmus obliquus*

```
  AAs = FFLLSS*SYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = -----------------------------------M----------------------------
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

23 – Code mitochondrial de *Thraustochytrium*

```
  AAs = FF*LSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ------------------------------M--M---------------M------------
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

24 – Code mitochondrial des ptérobranches

```
  AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSSKVVVVAAAADDEEGGGG
Starts = ---M---------------M---------------M---------------M------------
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

25 – Code des bactéries de la Candidate Division SR1 et des bactéries *Gracilis*

```
  AAs = FFLLSSSSYY**CCGWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M---------------M---------------M-----------M
                   M               M
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

26 – Code génétique nucléaire du champignon *Pachysolen tannophilus*

```
  AAs = FFLLSSSSYY**CC*WLLLAPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ------------------M---------------M----------------------------
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

27 – Code génétique nucléaire des ciliés de la classe *Karyorelictea*

```
  AAs = 
FFLLSSSSYYQQCCWWLLLAPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGGFFLLSSSSYYQQCCW
WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---------------*-------------------M----------------------------
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

28 – Code génétique nucléaire des condylostomes

```
  AAs = 
FFLLSSSSYYQQCCWWLLLAPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGGFFLLSSSSYYQQCCW
WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = -----------**--*-------------------M----------------------------
 Base1 = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
 Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
 Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

29 – Code génétique nucléaire des ciliés du genre *Mesodinium*

```
   AAs   =
FFLLSSSSYYYYCC*WLLLAPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGGWLLLLPPPPHHQQRR
RRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - M- - - - - - - - - - - - - - - - - - - - - - - - -
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

30 – Code génétique nucléaire des péritriches

```
   AAs   =
FFLLSSSSYYEECC*WLLLAPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGGWLLLLPPPPHHQQRR
RRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - M- - - - - - - - - - - - - - - - - - - - - - - - -
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
```

31 – Code génétique nucléaire du genre *Blastocrithidia*

```
   AAs   = FFLLSSSSYYEECCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = - - - - - - - - - - **- - - - - - - - - - - - - - - - - - - - - M- - - - - - - - - - - - - - - - - - - - - - - - -
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

33 - Code mitochondrial UAA-Tyr des vers marins *Cephalodiscidae*

```
   AAs   = FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSSKVVVVAAAADDEEGGGG
Starts = - - -M- - - - - - - - *- - - - - - - - M- - - - - - - - - - - - - -M- - - - - - - - - - - - - - -M- - - - - - - - - - -
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaagggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaagggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

[Fin de l'annexe (Norme ST.26 – Annexe I) et du document]