

## 非公式翻訳

### 非公式翻訳

本文書は、WIPO Standard ST.26の非公式の和訳文です。

WIPO国際事務局は、この文書の内容について責任を負わず、文書の正確性は検証していません。  
検証済みの内容については、[英語](#)、[スペイン語](#)、または[フランス語](#)の、WIPO Standard ST.26の公式文書をご参照ください。  
和訳文の提供元は日本の特許庁です。

### UNOFFICIAL TRANSLATION

This document is an unofficial translation of WIPO Standard ST.26 in Japanese language.

The International Bureau of WIPO is not responsible for this document and has not verified its accuracy. Refer to the official versions of WIPO  
Standard ST.26 in [English](#), [Spanish](#), or [French](#) for verified contents.  
Japanese Translation provided by the Japanese Patent Office (JPO).

## WIPO 標準 ST.26

XML（拡張可能なマークアップ言語）形式で  
塩基配列およびアミノ酸配列の配列表を表示するための WIPO 推奨標準

*Version 1.7*

2023 年 12 月 8 日開催の第 11 回 WIPO 標準委員会(CWS)会合により採択

WIPO 国際事務局により編修された文書

第 11 回 WIPO 標準委員会 (CWS) 会合は、WIPO 標準 ST.26 のバージョン 1.7 が 2024 年 7 月 1 日に発効することを決定した。  
その間、WIPO 標準 ST. 26 のバージョン 1.6 は引き続き有効であるものとする。

WIPO 標準は、産業財産権を所管する機関およびその他の関係者に情報を提供する目的で公開されている。

## 目次

はじめに .....	4
用語の定義 .....	4
範囲 .....	6
参考資料 .....	6
配列の表示 .....	7
塩基配列 .....	7
アミノ酸配列 .....	9
特殊な状況の表示 .....	11
XML 形式の配列表の構造 .....	12
ルート要素 .....	13
一般情報部 .....	13
配列データ部 .....	16
feature table .....	18
feature keys .....	18
必須の feature keys .....	19
feature location .....	19
feature qualifiers .....	21
必須の feature qualifiers .....	21
qualifier の要素 .....	21
フリーテキスト .....	24
コード配列 .....	25
変異体 .....	26

## 附属書

附属書 I 統制語彙

附属書 II 配列表の文書型定義(DTD)

附属書 III 配列表の実例(XML ファイル)

附属書 IV 配列表の XML インスタンスに用いる Unicode 基本ラテンコード表の文字サブセット

※附属書 V ~ VIIについては、WIPO 標準 ST.26 の公式版を御参照ください。

**WIPO 標準 ST.26**

XML（拡張可能なマークアップ言語）形式で

塩基配列表およびアミノ酸配列表を提示するための推奨標準

*Version 1.7*

2023年12月8日開催の第11回WIPO標準委員会(CWS)会合で採択

はじめに

1. 本標準は、特許出願において配列表への記載を要する塩基配列およびアミノ酸配列の開示情報、それらの開示情報の表示方法、XML形式での配列表の文書型の定義(DTD)を定義している。知的財産権を所管する機関は、特許出願の一部として、または特許出願に関連して提出された本標準に準拠型の配列表を受理することを推奨する。
2. 本標準の目的は次のとおりである。
  - (a) 特許出願人が、諸外国および国内または広域的な特許出願手続きの基準を満たす单一の配列表を作成できる様にすること。
  - (b) 特許出願人、一般公衆および審査官に利益をもたらし、普及を容易にするため、配列の表示の精度と品質を向上させること。
  - (c) 配列データの検索を容易にすること。および、
  - (d) 配列データを電子的な形式で交換し、コンピューター化されたデータベース上で配列データを利用できるようすること。

用語の定義

3. 本標準では、次の表現を以下の通り定義する。
  - (a) 「アミノ酸」とは、附属書I(第3節、表3)に記載されている記号のいずれかを用いて表示できる任意のアミノ酸を意味する。その様なアミノ酸には、特にD-アミノ酸および、修飾を受けた、または合成された側鎖を含むアミノ酸が含まれる。また本標準の第30項に従ってfeature tableに修飾を受けたものとして更に説明がない限り、アミノ酸とは非修飾のL-アミノ酸として解釈される。また本標準ではペプチド核酸(PNA)の残基はアミノ酸であるとみなされず、第3項(g)(i)(2)で定められている通りヌクレオチドであるとみなされる。
  - (b) 「統制語彙」とは本標準で使用されている用語であり、配列のfeature(配列の特徴)に関して記載する際に用いられなければならない用語である。すなわち、附属書Iで定められている通り、対象の領域や部位の注釈に用いられる用語である。
  - (c) 「残基の数え上げ」とは、特許出願において、各残基配列を順に数え上げて開示することを意味し、ここでは以下のようになる。
    - (i) 残基が名称、略語、記号または構造で示されている(例. HHHHHHQ または HisHisHisHisHisGln)。
    - または、
    - (ii) 複数の残基が簡略な式で表されている(例. His<sub>6</sub>Gln)。
  - (d) 「意図的にスキップされた配列」は空の配列としても知られており、出願開示内容との一貫性を保つために、配列表上の配列の番号付けを維持するための記入子を指し、例えば、開示内容と配列表の両方で配列の番号の振り直しを避けるため、開示内容から配列が削除される場合等に使用される。
  - (e) 「修飾アミノ酸」とは、第3項(a)で説明されている、L-アラニン、L-アルギニン、L-アスパラギ

ン、L-アスパラギン酸、L-システイン、L-グルタミン、L-グルタミン酸、L-グリシン、L-ヒスチジン、L-イソロイシン、L-ロイシン、L-リシン、L-メチオニン、L-フェニルアラニン、L-プロリン、L-ピロリシン、L-セリン、L-セレノシステイン、L-トレオニン、L-トリプトファン、L-チロシンまたはL-バリン以外のアミノ酸を意味する。

(f) 「修飾塩基」とは、デオキシアデノシン 5'-リン酸塩、デオキシグアノシン 5'-リン酸塩、デオキシシチジン 5'-リン酸塩、デオキシチミジン 5'-リン酸塩、アデノシン 5'-リン酸塩、グアノシン 5'-リン酸塩、シチジン 5'-リン酸塩またはウリジン 5'-リン酸塩以外の第 3(g)項に記載されたいずれかのヌクレオチドのことである。

(g) 「ヌクレオチド」とは、ヌクレオチドまたはヌクレオチド類似体に関する附属書 I (第 1 節、表 1) で定められている記号のいずれかを用いて表示できる、任意のヌクレオチドまたはヌクレオチド類似体のことである。

(i) 骨格部分は以下のいずれかから選択する。

(1) 2'デオキシリボース 5'リン酸 (デオキシリボヌクレオチドの主鎖の一部) または、リボース 5'リン酸 (リボヌクレオチドの主鎖の一部)。または、

(2) 核酸類似体の主鎖を形成するときに、2'デオキシリボース 5'リン酸またはリボース 5'リン酸の主鎖を含む核酸における核酸塩基の配列を模倣する核酸の配列をもたらす、2'デオキシリボース 5'リン酸またはリボース 5'リン酸の類似体であって、ここで、核酸類似体は相補的な核酸と塩基対をなすことができるものである。骨格部分の例として、ペプチド核酸のようなアミノ酸、グリコール核酸のようなグリコール分子、スレオース核酸のようなスレオフランオシル糖分子、モルホリノのようなモルホリン環およびホスホジアミデート基、およびシクロヘキセニル核酸のようなシクロヘキセニル分子などがある。

また、

(ii) 骨格部分は以下のいずれかとなる。

(1) 修飾を受けたまたは合成されたプリンまたはピリミジン核酸塩基を含む、核酸塩基と結合したもの。

または、

(2) ヌクレオチドが「AP 部位」または「脱塩基部位」と呼ばれる塩基配列の一部である場合、プリンまたはピリミジン核酸塩基を欠いているもの。

(h) 「残基」とは、配列内の任意の個々のヌクレオチドまたはアミノ酸、またはこれらの類似体を意味する。

(i) 「配列番号」とは、配列表の各配列に割り当てられた、その配列固有の番号（整数）を意味する。

(j) 「配列表」とは、出願時に提出される特許出願明細書の一部、または出願後に提出される書類であって、本標準に規定されている開示されたヌクレオチドおよび/またはアミノ酸の配列を、詳細な説明とともに包含するものを意味する。

(k) 「具体的に定義された」とは、記号「n」で表されるヌクレオチド以外の任意のヌクレオチドおよび記号「X」で表されるアミノ酸以外の任意のアミノ酸であって、附属書 I に記載されているものを意味する（それぞれ、第 1 項表 1 および第 3 項表 3 を参照のこと）。

(l) 「不明な」ヌクレオチドまたはアミノ酸とは、単一のヌクレオチドあるいはアミノ酸が存在するが、その同一性が不明であるか、または開示されていないものをいう。

(m) 「変異体配列」とは、一次配列に関して 1 つまたは複数の差異を含む塩基配列またはアミノ酸配列を意味する。これらの違いには、代替残基（第 15 項および第 27 項を参照）、変更残基（第 3 項 (g)、第 3 項 (h)、第 16 項、および第 29 項を参照）、欠失、挿入、および置換が含まれる場合がある。第 93 項から第 95 項参照のこと。

(n) 「フリーテキスト」は、特定の qualifier 値の形式の一種であり、説明文のフレーズまたはその他の指定された形式（附属書 I に示されている）の形式で表示される。第 85 項参照のこと。

(o) 「言語依存のフリーテキスト」とは、特定の *qualifier* のフリーテキスト値を意味する。これには、国際的、国内的、または広域的な手順の翻訳が必要になる場合がある。第 87 項参照のこと。

#### 4. 本標準では次の言葉を以下の通り定義する。

- (a) 法助動詞の「may」は、可能または許容を意味するが、必須ではない。
- (b) 法助動詞の「must」は本標準の定める要件であり、この要件を満たさない場合は不適合となる。
- (c) 法助動詞の「must not」は、本標準の禁止事項を表している。
- (d) 法助動詞の「should」は、強い推奨を意味するが、必須ではない。
- (e) 法助動詞の「should not」は、強い制止を意味するが、禁止ではない。

#### 範囲

5. 本標準は、特許出願において開示されている配列のヌクレオチドおよびアミノ酸の配列表を提示するための要件を定めるものである。

6. 本標準に準拠する配列表（以降、「配列表」という）は一般情報部と、配列データ部とを含んでいる。配列表は、附属書Ⅱに示されている文書型定義(DTD)を用いた XML 形式の單一ファイルとして提示しなければならない。一般情報部に含まれる書誌情報の目的は、配列表と、その配列表が提出された特許出願とを紐づけることのみにある。配列データ部は、1つまたは複数の配列データ要素で構成され、それぞれが 1 の配列に関する情報を含む。配列データ要素は、国際塩基配列データベースコラボレーション(INSDC)および Uniprot (アミノ酸配列とその機能情報を掲載しているデータベース) 仕様に基づく様々な feature keys とその後続の qualifiers を含むものである。

7. 本標準では、配列表への記載を要する配列とは、その残基の数え上げによって出願のどこにでも開示され、次のように表示できる配列である。

- (a) 10 個以上の具体的に定義されたヌクレオチドを含む非分岐配列または分岐配列の線形領域で、次の方法で隣接するヌクレオチドと結合する。
  - (i) 3'位から 5'位の間（または 5'位から 3'位の間）のホスホジエステル結合。  
または
  - (ii) 天然に存在する核酸中の核酸塩基の配置を模倣する隣接する核酸塩基の配置をもたらす化学結合。
- (b) 4 個以上の具体的に定義されたアミノ酸を含む非分枝配列または分岐配列の線形領域であって、その 4 つ以上の具体的に定義されたアミノ酸が結合して単一のペプチド主鎖を形成する（すなわち、隣接するアミノ酸がペプチド結合により結合する）。
- 8. 配列表には、個別の配列番号が付与された配列として、具体的に定義された 10 個未満のヌクレオチドまたは、具体的に定義された 4 個未満のアミノ酸の配列を記載してはならない。

#### 参考資料

9. 以下の標準および利用可能な情報に関する資料は、本標準と関連性がある。

国際塩基配列データベースコラボレーション(INSDC)  
国際標準化機構(ISO)言語コード 639-1: 2002 年公表  
Unitprot (蛋白質に関する共通情報)コンソーシアム  
W3C XML 1.0

<http://www.insdc.org/>

言語名を表示するためのコード – パート 1 : アルファベット 2 文字のコード

<http://www.uniprot.org/>

<http://www.w3.org/>

WIPO Standard ST.2  
WIPO Standard ST.3

グレゴリオ暦を使用してカレンダーの日付を指定する方法  
国、その他の団体および政府間組織を 2 文字で表すための推奨規格

WIPO Standard ST.16

異なる種類の特許文献を識別するための推奨標準コード

## 配列の表示

- 1 0. より長い配列の領域と同一の配列を含め、第 7 項に記載の各配列には、個別の配列番号を割り振らなければならない。配列番号は「1」から始まり、以降、整数の連番で増加する必要がある。配列識別番号は存在するがその番号に該当する配列が存在しない場合、すなわち意図的にスキップされた配列では、その配列の代わりに「000」を用いなければならない（第 58 項参照）。配列の総数は、配列が続くか、「000」が続くかにかかわらず、配列番号の総数と等しくならなければならない。

## 塩基配列

- 1 1. 塩基配列は、左から右への 5'位から 3'位方向、または 5'位から 3'位方向を模した左から右への一本鎖によってのみ表されなければならない。5'位および 3'位または他の同様の指定を配列に記載してはならない。両鎖の残基の数え上げによって開示された二本鎖のヌクレオチド配列は次のように表示しなければならない。
- (a) 2 つの別々の鎖が互いに完全に相補的である場合、単一の配列、または 2 つの別々の配列としてそれぞれに個別の配列番号が割り当てられる。または、
  - (b) 2 つの鎖が互いに完全には相補的でない場合、2 つの別々の配列であり、それぞれに個別の配列番号が割り当てられる。
- 1 2. 本標準では、配列に示される最初のヌクレオチドは残基位置番号 1 である。塩基配列が環状構造である場合、出願人は残基位置番号 1 のヌクレオチドを選択しなければならない。配列番号の割り振りは、5'末端から 3'末端の方向、または 5'末端から 3'末端の方向を模した方向に、配列全体に連番で割り振る。最後の残基位置番号は、配列内のヌクレオチド数と等しくなければならない。
- 1 3. 配列内のすべてのヌクレオチドは附属書 I（第 1 節、表 1 を参照のこと）で定められている記号を用いて表示しなければならない。その際は、アルファベットの小文字のみを用いて表示しなければならない。ヌクレオチドの表示に用いる記号は、1 つの残基のみを示す。
- 1 4. 記号「t」は、DNA ではチミン、RNA ではウラシルと解釈される。DNA のウラシルや RNA のチミンは、修飾塩基とみなされ、第 19 項で規定された feature table でさらに説明しなければならない。
- 1 5. 曖昧さ記号（2 つ以上の代替ヌクレオチドを表す）が適切な場合は、附属書 I（第 1 節、表 1）に記載されているように、最も限定的な記号を使用すべきである。例えば、特定の位置のヌクレオチドが「a」または「g」である可能性がある場合は、「n」ではなく「r」を使用すべきである。記号「n」は、feature table でさらに説明する場合を除いて、「a」、「c」、「g」または「t/u」のいずれかの 1 つであると解釈される。記号「n」は、ヌクレオチド以外のものを表すために用いてはならない。単一の修飾を受けた、または「不明な」ヌクレオチドは、第 16 項、17 項、21 項、または第 93 項から第 96 項に記載されているように、feature table におけるさらなる説明とともに記号「n」で表すことができる。配列変化、すなわち代替、欠失、挿入または置換の表現については、第 93 項から第 100 項を参照のこと。
- 1 6. 修飾塩基は、可能な限り対応する非修飾塩基のヌクレオチド、すなわち「a」、「c」、「g」または「t」として配列内で表現すべきである。配列内の修飾塩基のうち、附属書 I（第 1 節、表 1 参照）の他の記号で表すことができないもの、すなわち天然由来でないヌクレオチド等の「他の」ヌクレオチドは、記号「n」で表さなければならない。記号「n」は 1 つの残基のみを示す。
- 1 7. 修飾塩基に関しては、feature key 「modified\_base」およびその必須 qualifier 「mod\_base」を使用して、qualifier の値として附属書 I（第 2 節、表 2 参照）の略語の 1 つと組み合わせて、feature table（第 60 項以降参照）に詳しく説明しなければならない。略語が「OTHER」である場合は、その修飾塩基の完全で、省略されていない名称を、「note」qualifier の値として指定しなければならない。代替の修飾塩基のリストについては、qualifier 値「OTHER」をさらに「note」qualifier と組み合わせて使用しても良い（第 97 および 98 項参照）。前述の附属書 I（第 2 節、表 2 参照）に記載されている略語（または完全名称）は、配列自体で使用してはならない。
- 1 8. 同じ骨格部分を共有する連続した修飾塩基の 1 つまたは複数の領域を含む塩基配列（第 3 項(g)(i)(2)参

照) は、第 17 項の規定のとおり、さらに詳しく feature table で説明しなければならない。また第 22 項で規定されているように、領域内のすべての修飾塩基は、単一の INSDFeature 要素でまとめて記載しても良い。領域内のすべての修飾塩基を包含する最も限定的で省略されていない化学名または領域内のすべてのヌクレオチドの化学名のリストは、「note」qualifier の値として提供しなければならない。例えば、「a」、「c」、「g」または「t」の核酸塩基を含むグリコール核酸配列は、「note」qualifier に、「2,3-ジヒドロキシプロピルヌクレオシド」として記載し得る。あるいは、同じ配列を、「note」qualifier に、「2,3-ジヒドロキシプロピルアデニン、2,3-ジヒドロキシプロピルチミン、2,3-ジヒドロキシプロピルグアニン、2,3-ジヒドロキシプロピルシトシン」として記載し得る。領域内の個々の修飾塩基が追加の修飾を含む場合、その修飾塩基も、第 17 項に規定されているように、feature table に詳しく説明しなければならない。

19. DNA のウラシルやRNA のチミンは修飾塩基であるとみなされる。配列上では、「t」と表記されなければならない。さらに、feature table では、feature key 「modified\_base」、qualifier の値として qualifier 「mod\_base」に「OTHER」を組み合わせて、および、qualifier 値として qualifier 「note」にそれぞれ「uracil」または「thymine」を組み合わせて表記されなければならない。

20. 次の例は、上記第 16 項から第 18 項による修飾塩基の表示例を示す。

例 1：附属書 I の略語を用いた修飾塩基（第 2 節、表 2 参照）

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>15</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例 2：附属書 I の「OTHER」を用いた修飾塩基（第 2 節、表 2 参照）

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>4</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例 3：第 3 項(g)(i)(2)に包含される修飾塩基からなる塩基配列であって、さらに修飾を受けた 2 つの個別のヌクレオチドを有するもの

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>1..954</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
```

```

<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>2,3-dihydroxypropyl nucleosides</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_quals>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>439</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>i</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>684</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>xanthine</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>

```

- 2 1. 「不明な」ヌクレオチドはすべて、配列内の記号「n」で表さなければならない。また「不明な」ヌクレオチドは feature key 「unsure」を使用して、feature table (第 60 項以降参照)でさらに詳しく説明すべきである。記号「n」は、1つの残基のみを示す。
- 2 2. 同じ記載が適用される既知数の連続した「a」、「c」、「g」、「t」または「n」 残基を含む領域は、要素 INSDFeature\_location に存在位置記述子として、「x...y」という構文を持つ単一の要素 INSDFeature element を用いてまとめて記載し得る (第 64 項から第 71 項参照)。配列変化、すなわち、代替、欠失、挿入または置換の表現については、第 93 項から第 100 項を参照のこと。
- 2 3. 次の例は、上記第 22 項に従い、同一の記載が適用される修飾塩基の領域の表現を示している。

```

<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>358..485</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>isoguanine</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>

```

## アミノ酸配列

- 2 4. アミノ酸配列のアミノ酸は、左から右に向かって、アミノ基からカルボキシ基の方向で表されなければならない。アミノ基およびカルボキシ基は配列内に表示してはならない。

- 2 5. 本標準では、配列の最初のアミノ酸は残基番号 1 であり、成熟蛋白質に先行するアミノ酸、例えば、プレ配列、プロ配列、プレプロ配列、シグナル配列なども含まれる。アミノ酸配列が環状で、その環がペプチド結合で結合されたアミノ酸残基のみからなる場合、すなわち、配列にアミノ末端およびカルボキシ末端が無い場合、出願人は残基位置 1 のアミノ酸を選択しなければならない。番号の割り振りは、配列全体を通して、アミノ基からカルボキシ基の方向に連続する。
- 2 6. 配列内のすべてのアミノ酸は、附属書 I（第 3 節、表 3 参照）に記載されている記号を用いて表示しなければならない。アルファベットの大文字のみを使用しなければならない。アミノ酸の表示に用いる記号は、1 つの残基のみを示す。
- 2 7. 曖昧さ記号（代替となる 2 以上のアミノ酸を表す）が適切である場合には、附属書 I（第 3 章、表 3）に記載されているように、最も限定的な記号を使用すべきである。例えば、特定の位置のアミノ酸がアスパラギン酸またはアスパラギンである可能性がある場合には、「X」ではなく「B」という記号を使用すべきである。記号「X」は、feature table の詳細な情報とともに使用される場合を除き、「A」、「R」、「N」、「D」、「C」、「Q」、「E」、「G」、「H」、「I」、「L」、「K」、「M」、「F」、「P」、「O」、「S」、「U」、「T」、「W」、「Y」、または「V」のいずれか 1 つと解釈される。記号「X」は、アミノ酸以外のものを表すために使用してはならない。単一の修飾アミノ酸または「不明な」アミノ酸は、例えば、第 29 項、第 30 項、第 32 項、または第 92 項から第 98 項で提供されるような feature table に詳しい説明とともに、記号「X」で表すことができる。配列変化、すなわち、代替、欠失、挿入または置換の表現については、第 93 項から第 100 項を参照のこと。
- 2 8. 例えば、「Ter」、アスタリスク「\*」、ピリオド「.」または空白「 」で表される内部終止記号で区切られた開示されたアミノ酸配列は、少なくとも 4 つの具体的に定義されたアミノ酸を含み、かつ第 7 項に包含される各アミノ酸配列について、個別の配列として含めなければならない。このような個別の配列には、それぞれ独自の配列番号を割り当てなければならない。終止記号および空白は、配列表の配列に含めてはならない（第 57 項参照）。
- 2 9. D-アミノ酸を含む修飾アミノ酸は、可能な限り、対応する非修飾アミノ酸として配列に表示すべきである。配列の修飾アミノ酸のうち、附属書 I（第 3 節、表 3 参照）の他の記号で表すことができないもの、すなわち、「その他」のアミノ酸は、「X」で表示しなければならない。記号「X」は、1 つの残基のみを示す。
- 3 0. 修飾アミノ酸は、feature table にさらに記述しなければならない（第 60 項以降を参照のこと）。必要に応じて、feature key 「CARBOHYD」または「LIPID」は、「note」qualifier と共に使用すべきである。他の翻訳後の修飾アミノ酸については、「note」qualifier と共に feature key 「MOD\_RES」を使用すべきであり、それ以外の場合は、feature key 「SITE」を「note」qualifier とともに使用する必要がある。「note」qualifier の値は、附属書 I に記載されている略語（第 4 節、表 4 参照）または修飾アミノ酸の省略されていない完全な名称のいずれかでなければならない。上記の表 4 に記載されている略語や、省略されていない完全な名称は、配列自体に使用してはならない。

- 3 1. 次の例は、上記第 30 項による修飾アミノ酸の表示である。

例 1：翻訳後修飾アミノ酸

```
<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>3Hyp</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例 2：翻訳後非修飾アミノ酸

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
```

```
<INSDFeature_location>3</INSDFeature_location>
<INSDFeature_quals>
  <INSDQualifier>
    <INSDQualifier_name>note</INSDQualifier_name>
    <INSDQualifier_value>Orn</INSDQualifier_value>
  </INSDQualifier>
</INSDFeature_quals>
</INSDFeature>
```

### 例 3 : D-アミノ酸

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>9</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>D-Arginine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

3 2. 「不明な」アミノ酸は、配列に記号「X」を用いて表示しなければならない。「X」と指定された「不明な」アミノ酸は、feature key 「UNSURE」と、任意で「note」qualifier を用いて、feature table（第 60 項以降参照）にさらに記述しなければならない。記号「X」は、1 つの残基のみを示す。

3 3. 次の例は、上記第 32 項による「不明な」アミノ酸の表示である。

```
<INSDFeature>
  <INSDFeature_key>UNSURE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>A or V</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

3 4. 同じ記述が適用される既知数の連続した「X」残基を含む領域は、要素 INSDFeature\_location のロケーション記述子として「x..y」という構文を用いてまとめて記載しても良い(第 64 項から 70 項参照)。配列変化、すなわち、代替、欠失、挿入、または置換の表現については、第 93 項から 100 項を参照のこと。

### 特殊な状況の表示

3 5. 残基の数え上げによって開示された配列で、より長い配列の 1 つまたは複数の非連続セグメント、または異なる配列のセグメントから单一の連続配列として構成されているものは、配列表に記載しなければならず、個別の配列番号を割り当てなければならない。

3 6. 具体的に定義された残基の領域が、連続した「n」または「X」残基の 1 つまたは複数の領域によって分離されている配列（それぞれ第 15 項および第 27 項参照）で、各領域の「n」または「X」残基の正確な数が開示されているものは、単一の配列として配列表に記載しなければならず、個別の配列番号を割り当てなければならない。

3 7. 具体的に定義された残基の領域が、不明のまたは未公表の数の残基の 1 つまたは複数のギャップによって分離されている配列は、単一の配列として配列表に記載してはならない。第 7 項に記載されている特異的に定義された残基の各領域は、個別の配列として配列表に記載しなければならず、個別の識別番号が割り当てられなければならない。

## XML 形式の配列表の構造

3 8. 上記第 6 項に従い、本標準に準拠型の配列表ファイルの XML インスタンスは次のとおりである。

- (a) 一般情報部：配列表の対象となる特許出願に関する情報を含み、および、
- (b) 配列データ部：1 つまたは複数の配列データ要素を包み、各要素には 1 つの配列に関する情報を含む。

配列表の例は、附属書Ⅲに記載されている。

3 9. 配列表は、附属書Ⅱ「配列表の文書型定義(DTD)」に示す DTD を用いて、XML (ver1.0) で提示しなければならない。

- (a) XML インスタンスの 1 行目には、XML 宣言を記載しなければならない。

<?xml version="1.0" encoding="UTF-8"?>

- (b) XML インスタンスの 2 行目には、文書型(DOCTYPE)宣言が含まれていなければならない。

<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3 //EN"  
"ST26SequenceListing\_V1\_3.dtd">.

4 0. 電子配列表全体を 1 つのファイルに含めなければならない。ファイルは Unicode UTF-8 を使用してエンコードしなければならないが、次の制限がある。

- (a) 一般情報部の要素である ApplicantName、InventorName、InventionTitle、および配列データ部の NonEnglishQualifier\_value に含まれる情報は、Unicode Control コードポイント 0000-001F および 007F-009F を除き、XML 1.0 仕様で示される有効な Unicode 文字で構成し得る。  
予約文字「」、「&」、「」、「<」、「>」（それぞれ、Unicode コードポイント 0022、0026、0027、003C、003E）は、第 41 項に記載されているように置き換えなければならない。そして、
- (b) 一般情報部の他のすべての要素および属性、ならびに配列データ部の他のすべての要素および属性に含まれる情報は、Unicode 基本ラテンコード表の印刷可能な文字（スペース文字を含む）で構成されなければならない（すなわち、Unicode コードポイント 0020 から 007E までに限定される- 附属書Ⅳ 参照）。予約文字「」、「&」、「」、「<」、「>」（それぞれ Unicode コードポイント 0022、0026、0027、003C、003E）は、第 41 項に記載されているように置き換えなければならない。

4 1. 配列表の XML インスタンスでは、数値文字参照<sup>1</sup>を使用してはならず、要素の属性または内容の値で使用する場合は、次の予約文字を対応する定義済み実体に置き換えなければならない。

予約文字	定義済み実体
<	&lt;
>	&gt;
&	&amp;
“	&quot;
‘	&apos;

許可される文字参照は、この項で規定されている定義済み文字実体のみである。例として、第 71 項を参照のこと。

4 2. すべての必須要素にデータを入力しなければならない（意図的にスキップされた配列について第 58 項で規定されている場合を除く）。内容が利用できない任意要素は、XML インスタンスに表示すべきではない（配列での「replace」qualifier の値の削除表現について第 97 項で規定されている場合を除く）。

<sup>1</sup> 数値文字参照は、そのユニバーサル文字セット/ユニコードコードポイントによって文字を参照し、「&#nnnn;」または「&#xhhhh;」の形式を使用する。ここで、「nnnn」は 10 進形式のコードポイントであり、

「hhhh」は16進形式のコードポイントである。

## ルート要素

4 3. 本標準に準拠型の XML インスタンスのルート要素は、次の属性を持つ要素 ST26SequenceListing である。

属性	記載情報	必須／任意
dtdVersion	「V# #」の形式でこのファイルを作成するために使用する DTD のバージョン、e.g., 「V1_3」。	必須
fileName	配列表ファイルの名称	任意
softwareName	ファイルの生成に用いたソフトウエア名	任意
softwareVersion	ファイルの生成に用いたソフトウエアのバージョン	任意
productionDate	配列表ファイルの作成日(ひな型は、「CCYY-MM-DD」)	任意
originalFreeTextLanguageCode	言語に依存するフリーテキスト qualifier が作成された単一の原語の言語コード（第 9 項の ISO 639-1:2002 への参照文を参照）	任意
nonEnglishFreeTextLanguageCode	要素 NonEnglishQualifier_value の言語コード（第 9 項の ISO 639-1:2002 への参照文を参照）	配列表に要素 NonEnglishQualifier_value が存在する場合は必須

4 4. 次の例は、上記第 43 項の XML インスタンスのルート要素 T26SequenceListing およびその属性を示している。

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11-405455-SEQL.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="de" nonEnglishFreeTextLanguageCode="fr">
{...}*
</ST26SequenceListing>
```

\*{...}\* は、この例に含まれていない一般情報部と配列データ部を示している。

## 一般情報部

4 5. 一般情報部の要素は、次のように特許出願情報に関連している。

要素	記載情報	必須／任意
ApplicationIdentification	特許出願番号（配列表はこの特許出願番号で提出される）	出願番号の割り当て後、配列表を提出する場合は必須
The ApplicationIdentification (以下で構成される)		
IPOfficeCode	特許出願した知財庁の ST.3 コード	必須
ApplicationNumberText	特許出願した知財庁が割り当てた特許出願番号(例. PCT/IB2013/099999)	必須
FilingDate	配列表が提出された特許出願の出願日(ST.2 のひな型 “CCYY-MM-DD”、すなわち暦年(4 衢)、暦月(2 衢)および暦月の暦日(2 衢)、例. 2015-01-31)	出願日の確定後は、配列表を提出する場合は必須

ApplicantFileReference	第 40 項(b)に記載のある文字で入力された、出願人が特定の出願を識別するために割り当てる個別の識別番号	出願番号の割り当て前の時点で配列表を提出する場合は必須、それ以外は任意
EarliestPriorityApplicationIdentification	最先の優先権出願の識別情報 (IPOfficeCode、ApplicationNumberText および FilingDate も含む。上記 ApplicationIdentification 参照)	優先権が主張されている場合は必須
ApplicantName	第 40 項(a)に記載のある文字で入力された最初に挙げられた出願人の名前。この要素には、第 47 項に規定する必須属性 languageCode が含まれる。	必須
ApplicantNameLatin	ApplicantName が第 40 項(b)に記載のある文字以外で入力されている場合、最初に挙げられた出願人名の翻訳または翻字は第 40 項(b)に定める文字でも入力しなければならない。	出願人名にラテン文字以外の文字が含まれている場合は必須
InventorName	第 40 項(a)に記載のある文字で入力された最初に挙げられた発明者名。この要素には、第 47 項にきさいのある必須属性 languageCode が含まれる。	任意
InventorNameLatin	InventorName が第 40 項(b)に記載のある文字以外で入力されている場合、最初に挙げられた発明者の翻訳または翻字は第 40 項(b)に定める文字で入力しても良い。	任意
InventionTitle	出願言語で、第 40 項(a)に記載のある文字で入力された発明の名称。追加言語への発明の名称の翻訳は、要素 Invention Title を使用して、第 40 項(a)に記載のある文字で入力しても良い。この要素には、第 48 項に記載のある必須属性 languageCode を含む。発明の名称は 2 文字から 7 文字とされるべきである。	出願言語で記入する場合は必須、追加言語で記入する場合は任意
SequenceTotalQuantity	意図的にスキップされた配列（空の配列とも呼ばれる）を含む、配列表の全配列の総数（第 10 項参照）。	必須

4.6. 次の例は、上記の第 45 項に準拠型の配列表の一般情報部の表示を示す。

例 1：出願識別情報および出願日が付与される前に提出された配列表

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" 14ilename="Invention_SEQL.xml" softwareName="WIPO
Sequence" softwareVersion="1.0" productionDate="2022-05-10"
```

```

originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
    <ApplicantFileReference>AB123</ApplicantFileReference>
    <EarliestPriorityApplicationIdentification>
        <IPOfficeCode>IB</IPOfficeCode>
        <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
    </EarliestPriorityApplicationIdentification>
    <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
    <InventorName languageCode="en">Keiko Nakamura</InventorName>
    <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
    <SequenceTotalQuantity>9</SequenceTotalQuantity>
    <SequenceData sequenceIDNumber="1"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="2"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="3"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="4"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="5"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="6"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="7"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="8"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="9"> {...}</SequenceData>
</ST26SequenceListing>

```

\*{...} は、この例に含まれていない各配列の関連情報を表している。

例 2：出願識別番号および出願日が付与された後に提出された配列表

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="1_3" 15ilename="Invention_SEQL.xml" softwareName="WIPO Sequence"
softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
    <ApplicationIdentification>
        <IPOfficeCode>US</IPOfficeCode>
        <ApplicationNumberText>14/999,999</ApplicationNumberText>
        <FilingDate> 2015-01-05</FilingDate>
    </ApplicationIdentification>
    <ApplicantFileReference>AB123</ApplicantFileReference>
    <EarliestPriorityApplicationIdentification>
        <IPOfficeCode>IB</IPOfficeCode>
        <ApplicationNumberText>PCT/IB2014/099999</ApplicationNumberText>
        <FilingDate>2014-07-10</FilingDate>
    </EarliestPriorityApplicationIdentification>
    <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
    <InventorName languageCode="en">Keiko Nakamura</InventorName>
    <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
    <SequenceTotalQuantity>9</SequenceTotalQuantity>
    <SequenceData sequenceIDNumber="1"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="2"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="3"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="4"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="5"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="6"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="7"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="8"> {...}</SequenceData>
    <SequenceData sequenceIDNumber="9"> {...}</SequenceData>
</ST26SequenceListing>

```

\*{...} は、この例に含まれていない各配列の関連情報を表している。

4 7. 出願人名および、任意に発明者名は、通常、出願言語で参照されるため、要素 ApplicantName および InventorName でそれぞれ示さなければならない。適切な言語コード（第 9 項の「国際標準化機構(ISO)言語コード 639-1:2002 年公表」の参照を参照）は、各要素の属性 languageCode で示さなければならない。示された出願人名に、第 40 項(b)に記載のあるラテン文字以外の文字を含む場合、出願人名の翻字または翻訳は要素 ApplicantNameLatin で、ラテン文字でも示さなければならない。示された発明者名がラテン文字以外の文字を含む場合、発明者名の翻字または翻訳は、要素 InventorNameLatin で、ラテン文字で示しても良い。

4 8. 発明の名称は、要素 InventionTitle において出願言語で示さなければならず、複数の要素 InventionTitle を使用して追加言語で示しても良い（第 45 項の表を参照）。適切な言語コード（第 9 項の「国際標準化機構(ISO)言語コード 639-1:2002 年公表」の参照を参照）は、要素の属性 languageCode で示さなければならない。

4 9. 次の例は、上記の第 47 項および第 48 項による発明者名および発明の名称の表示を示す。

例：出願人名および発明者名をそれぞれ日本語とラテン文字で表記し、発明の名称を日本語、英語およびフランス語で表記する。

```
<ApplicantName languageCode="ja">出願製薬株式会社</ApplicantName>
<ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantNameLatin>
<InventorName languageCode="ja">特許 太郎</InventorName>
<InventorNameLatin>Taro Tokkyo</InventorNameLatin>
<InventionTitle languageCode="ja">efg 蛋白質をコードするマウス abcd-1 遺伝子</InventionTitle>
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
<InventionTitle languageCode="fr">Gène abcd-1 de Mus musculus pour protéine efg</InventionTitle>
```

#### 配列データ部

5 0. 配列データ部は、1 つまたは複数の要素 SequenceData で構成されていなければならず、各要素には 1 つの配列に関する情報が含まれていなければならない。

5 1. 各要素 SequenceData には必須要素 sequenceIDNumber を有し、各配列の配列番号（第 10 項参照）が含まれていなければならない。例えば、以下のようにになる。

```
<SequenceData sequenceIDNumber="1">
```

5 2. 要素 SequenceData は、次のような従属要素で構成される従属要素 INSDSeq を含まなければならない。

要素	記載	必須／任意	
		配列	意図的にスキップされた配列
INSDSeq_length	配列の長さ	必須	必須 値は不要
INSDSeq_moltype	分子の種類	必須	必須 値は不要
INSDSeq_division	特許出願と関連する配列であるという表示	「PAT」という値が付いている場合は、必須	必須 値は不要
INSDSeq_feature-table	配列の注釈を示すリスト	必須	含めてはならない
INSDSeq_sequence	配列	必須	必須 値に「000」を記述する。

5 3. 要素 INSDSeq\_length は、要素 INSDSeq\_sequence に含まれる配列のヌクレオチドまたはアミノ酸の数を開示しなければならない。例えば、次のようにになる。

```
<INSDSeq_length>8</INSDSeq_length>
```

5 4. 要素 INSDSeq\_moltype では、表現されている分子タイプを開示しなければならない。塩基配列（塩基類似体の配列を含む）の場合、分子タイプは DNA または RNA として示さなければならない。アミノ酸配列の場合、分子タイプは AA と表示しなければならない。（この要素は、第 55 項および第 84 項で説明している qualifier 「mol\_type」とは異なる）。例えば、次のようにする。

```
<INSDSeq_moltype>AA</INSDSeq_moltype>
```

5 5. 1つまたは複数のスクレオチドの DNA および RNA 両方のセグメントを含む塩基配列の場合、分子タイプはDNAと表示しなければならない。複合 DNA／RNA 分子は、feature key 「source」、必須 qualifier 「organism」の値が「synthetic construct」、必須 qualifier 「mol\_type」の値が「otherDNA」であることを用いて、feature table にさらに記述しなければならない。複合 DNA／RNA 分子の DNA および RNA セグメントは、feature key 「misc\_feature」およびその qualifier 「note」を用いてさらに記述しなければならない。

5 6. 次の例では、上記の第 55 項に準じて、DNA セグメントおよびRNA セグメント両方を含む塩基配列の表示を示す。

```
<INSDSeq>
  <INSDSeq_length>120</INSDSeq_length>
  <INSDSeq_moltype>DNA</INSDSeq_moltype>
  <INSDSeq_division>PAT</INSDSeq_division>
  <INSDSeq_feature-table>
    <INSDFeature>
      <INSDFeature_key>source</INSDFeature_key>
      <INSDFeature_location>1..120</INSDFeature_location>
      <INSDFeature_quals>
        <INSDQualifier>
          <INSDQualifier_name>organism</INSDQualifier_name>
          <INSDQualifier_value>synthetic construct</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
          <INSDQualifier_name>mol_type</INSDQualifier_name>
          <INSDQualifier_value>other DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_quals>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>1..60</INSDFeature_location>
      <INSDFeature_quals>
        <INSDQualifier>
          <INSDQualifier_name>note</INSDQualifier_name>
          <INSDQualifier_value>DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_quals>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>61..120</INSDFeature_location>
      <INSDFeature_quals>
        <INSDQualifier>
          <INSDQualifier_name>note</INSDQualifier_name>
          <INSDQualifier_value>RNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_quals>
    </INSDFeature>
  </INSDSeq_feature-table>
<INSDSeq_sequence>cgaccacgcgtccgaggaaccaaccatcacgttgaggactcgtgaaggaattggataatcccgccatccaaaatg
gcgagcgccgactcattgtcctgtaccgtcgagcggc</INSDSeq_sequence>
</INSDSeq>
```

- 5 7. 要素 INSDSeq\_sequence は、配列を開示しなければならない。附属書 I に記載されている適切な記号（第 1 節、表 1 および第 3 節、表 3）のみを配列に含めなければならない。数字、句読点または空白文字を配列に含めてはならない。
- 5 8. 意図的にスキップされた配列は、配列表に記載しなければならず、次のように表示しなければならない。
- 要素 SequenceData とその属性 sequenceIDNumber で、スキップされた配列の配列番号が値として提供される。
  - 要素 INSDSeq\_length, INSDSeq\_moltype, INSDSeq\_division は存在するが値は提供されない。
  - 要素 INSDSeq\_feature-table を記載してはならない。また、
  - 文字列 「000」を値とする要素 INSDSeq\_sequence

5 9. 次の例は、上記の第 58 項に従い、意図的にスキップされた配列の表示を示す。

```
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length/>
    <INSDSeq_moltype/>
    <INSDSeq_division/>
    <INSDSeq_sequence>000</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
```

feature table

6 0. Feature table には、特定の配列内の様々な領域の位置と役割に関する情報が含まれている。Feature table

は、意図的にスキップされた配列以外のすべての配列に必要である。Feature table は 1 つまたは複数の要素 INSDFeature からなる要素 INSDSeq\_feature-table に含まれる。

6 1. 各要素 INSDFeature は、1 つの feature を記述し、以下のような従属要素で構成されている。

要素	記載	必須／任意
INSDFeature_key	feature を示す文字または略語	必須
INSDFeature_location	feature に対応する配列内の領域	必須
INSDFeature_quals	feature に関する補助情報を含む qualifier	feature key が 1 つまたは複数の qualifier (例. Source, otherwise, Optional)を必要とする場合は必須

feature keys

6 2. 附属書 I には、本標準に使用されなければならない feature key の一覧と、関連する qualifier の一覧、及びこれらの qualifier が必須であるか任意であるかの指示が記載されている。附属書 I の第 5 節には塩基配

列に用いる feature key が、第 7 節にはアミノ酸配列に用いる feature key がそれぞれ掲載されている。

### 必須の feature keys

- 6 3. 意図的にスキップされた配列を除き、feature key 「source」はすべての塩基配列およびすべてのアミノ酸配列に必須である。各配列には、その配列全体をその範囲とする単一の feature key 「source」が必要である。1つの配列が複数の起源に由来する場合、それらの起源は、塩基配列では feature key 「misc\_feature」とその qualifier 「note」を、アミノ酸配列では feature key 「REGION」とその qualifier 「note」を用いて、feature table にさらに記述することができる。

### Feature location

- 6 4. 必須要素 INSDFeature\_location は、要素 INSDSeq\_sequence 内の feature に対応する部位または領域を定義する、少なくとも 1 つの存在位置記述子を含まなければならない。アミノ酸配列は、必須要素 INSDFeature\_location に 1 つのみ存在位置記述子を含まなければならない。塩基配列は、1 つまたは複数の存在位置演算子と組み合わせて使用される場合、必須要素 INSDFeature\_location に 1 つまたは複数の存在位置記述子を持つことができる（第 67 項から 70 項参照）。
- 6 5. 存在位置記述子は、単一の残基番号、残基番号の連続した範囲を区切る領域、または特定の残基または残基の範囲を超えて広がる部位または領域とすることができます。存在位置記述子は、要素 INSDSeq\_sequence の範囲を超える残基の数を含んではならない。塩基配列のみ、存在位置記述子は 2 つの隣接する残基番号間の部位とすることができます。ある配列の特徴が塩基配列の不連続な部位や領域に対応する場合には、複数の存在位置記述子を存在位置演算子と組み合わせて使用しなければならない（第 67 項から 70 項参照）。
- 6 6. 各タイプの存在位置記述子の構文を以下の表に示す。X と y は正の整数で示される残機番号で、要素 INSDSeq\_sequence の配列の長さ以下であり、x は y 未満である。

#### (a) 塩基およびアミノ酸配列の存在位置記述子

存在位置記述子のタイプ	構文	記載
単一の残基番号	X	配列上の単一の残基を指す。
配列の範囲を区切る残基番号	X..y	先頭の残基と末端の残基で囲まれた、残基の連続した範囲を指す。
特定の区間の先頭の残基番号の前に存在する残基と、特定の区間の末端の残基番号の後に存在する残基	<X >X <X..Y X..>Y <X..>Y	指定された残基または残基の範囲を含み、指定された残基を超えて広がる領域を指す。記号「<」および「>」は単一の残基、または残基の範囲の先頭および末端の残基番号とともに使用しても良く、feature が指定された残基番号を超えて広がることを示すことができる。

#### (b) 塩基配列のみの存在位置記述子

存在位置記述子のタイプ	構文	記載
2 つの隣接するヌクレオチド間に存在する部位	X^Y	2 つの隣接するヌクレオチド間の部位（エンドヌクレアーゼによる切断部位など）を指す。隣接するヌクレオチドの位置番号はカラット(^\)で区切られている。この記述子に許される形式は、X^X+1 (例えば 55^56)、または環状ヌクレオチドの場合は X^1 で、ここで「X」は分子の全長で、例えば長さ 1000 の環状分子の場合は 1000^1 となる。

#### (c) © アミノ酸配列のみの存在位置記述子

存在位置記述子のタイプ	構文	記載
鎖内架橋により結合した残基番号	X..Y	「CROSSLINK」や「DISULFID」のような鎖内架橋を示す機能と共に使用される場合、鎖内架橋によって結合されたアミノ酸を指す。

- 6 7. 塩基配列の要素 INSDFeature\_location には、1 つまたは複数の存在位置演算子が含まれることがある。存在位置演算子は、单一だが不連続な feature に対応する 1 つの存在位置記述子または存在位置記述子の組み合わせの接頭辞であり、示された配列上の feature に対応する位置情報、また

はその feature の構築方法を指定する。次に、存在位置演算子の一覧とその定義を示す。ロケーション演算子はヌクレオチドに対してのみ使用できる。

location の構文	location の記載
join(location,location, ... location)	示されている位置情報は、1つの連続した配列になる様に結合している(端-端接合)。
order(location,location, ... location)	要素は特定の順序で表示されているが、これらの要素をまとめて表示することが適切であるかどうかに関しては示唆されていない。
complement(location)	5' 位から 3' 位の方向、または 5' 位から 3' 位の方向を模した方向に読んだ場合、存在位置記述子で指定された配列範囲に相補的な鎖上に feature が位置することを示している。

- 6 8. join および order 存在位置演算子では、少なくとも 2 つのコンマ区切りの存在位置記述子を指定する必要がある。2つの隣接する残基間の部位を含む存在位置記述子 ( $x^y$ ) は、join または order location 内で使用してはならない。join 存在位置演算子の使用は、存在位置記述子によって記述された残基が生物学的プロセスによって物理的に接触していることを意味する（例えば、コーディング領域の feature に寄与するエクソンなど）。
- 6 9. 存在位置演算子「complement」は、同じロケーションで「join」または「order」と組み合わせて使用できる。同じロケーションで「join」と「order」を組み合わせて使用してはならない。

7 0. 次の例は、上記の第 64 項から 69 項に従って、feature location の表示を示す。

(d) 塩基配列およびアミノ酸配列の location

location の例	記載
467	配列内の残基 467 を示している。
340..565	残基 340 と残基 565 に囲まれた、残基の連続した範囲を示している。
<1	先頭の残基の前の feature location を示している。
<345..500	feature の正確な下側の境界点が不明であることを示している。その location は 345 の残基の前に位置するいずれかの残基から始まって、残基 500 の位置まで（残基 500 を含む）継続していることを示している。
<1..888	先頭に配列された残基の前から始まり、残基 888 まで（残基 888 を含む）継続する feature を示している。
1..>888	先頭に配列された残基から始まり、残基 888 を超えて継続する feature を示している。
<1..>888	先頭に配列された残基の前から始まり、残基 888 を超えて継続する feature を示している。

(b) 塩基配列のみに関するlocation

location の例	記載
123^124	残基 123 と残基 124 間の部位を示している。
join(12..78,134..202)	領域 12～78 と 134～202 を結合して、1つの連続した配列を形成することを示している。
complement(34..126)	ヌクレオチド 126 を補完するヌクレオチドを先頭に、ヌクレオチド 34 を補完するヌクレオチドを末端とすることを示している (feature は、残基番号が付いた本鎖を補完する鎖上に存在する)。
complement(join(2691..4571, 4918..5163))	ヌクレオチド 2691 からヌクレオチド 4571 までと、ヌクレオチド 4918 からヌクレオチド 5163 までが結合しており、feature はその結合したセグメントを補完していることを示している (feature はこの結合した鎖を補完する鎖上に存在する)。

join(complement(4918..5163), complement(2691..4571))	ヌクレオチド 4918 からヌクレオチド 5163 までの領域と、ヌクレオチド 2691 からヌクレオチド 4571 までの領域を補完しており、補完するセグメントと補完されるセグメントが結合していることを示している (feature はこの結合した鎖を補完する鎖上に存在する)。
--	---

## (c) アミノ酸配列のみに関するlocation

location の例	記載
340..565	「CROSSLNK」や「DISULFID」のような鎖内架橋を示す feature と共に使用すると、340 位と 565 位のアミノ酸が鎖内架橋で結合していることを示している。

7 1. 配列表の XML インスタンスでは、存在位置記述子の文字「<」および「>」は、適切な予め定義された文字実体に置き換えなければならない (第 4 1 項参照)。例えば、次のようにになる。

Feature location "<1":

```
<INSDFeature_location>&lt;1</INSDFeature_location>
```

Feature location "1..>888":

```
<INSDFeature_location>1..&gt;888</INSDFeature_location>
```

feature qualifiers

7 2. qualifier は、feature key および feature location で伝えられる情報に加えて、feature に関する情報を提供するために使用される。qualifier によって伝えられる様々な種類の情報に対応するために、次の 3 種類の値の形式がある。

(a) フリーテキスト (第 85 項から 87 項参照)

(b) 統制語彙または計数値 (例えば、数字や日付) および、

(c) 配列

7 3. 附属書 I の第 6 節では、各塩基配列の feature key に用いる qualifier とその指定された値の形式のリストが、第 8 節では、各アミノ酸配列の feature key に用いる qualifier と (もしあれば) その指定された値の形式のリストが示されている。

7 4. 第 7 項に含まれる配列のうち、qualifier の値として提供されるものは、配列表に別途記載され、独自の配列番号が割り振られなければならない (第 10 節参照)。

必須の feature qualifiers

7 5. 必須の feature key、すなわち、塩基配列およびアミノ酸配列の「source」には、「organism」および「mol\_type」という 2 つの必須 qualifier が必要である。一部の任意の feature key にも、必須 qualifier が必要である。

qualifier の要素

7 6. 要素 INSDFeature\_quals は、1 つまたは複数の要素 INSDQualifier を含む。各要素 INSDQualifier は 1 つの qualifier を示しており、次の 3 つの従属要素と 1 つの任意属性で構成されている。

要素／属性	記載されている情報	必須／任意
INSDQualifier_name	qualifier の名称 (附属書 I、第 6 節および第 8 節参照)。	必須

INSDQualifier_value	qualifier の値がある場合は、指定された形式（附属書 I、第 6 節および第 8 節参照）で、および、第 40 項(b)に記載された文字で構成される。	指定されている場合は必須（第 87 項、附属書 I、第 6 節および第 8 節参照）
NonEnglishQualifier_value	qualifier の値がある場合には、指定された形式（附属書 I の第 6 節および第 8 節参照）で、および、第 40 項(a)に記載された文字で構成される。	指定されている場合は必須（第 87 項、附属書 I、第 6 節および第 8 節参照）
id	言語に依存するフリーテキスト値を有する qualifier は、要素 INSDQualifier の任意の XML 属性「id」を用いることで一意に識別できる（第 87 項(d)参照）。属性「id」の値は文字「q」から始め、任意の正の整数で続けなければならない。属性「id」の値は、1 つの要素 INSDQualifier に対して一意でなければならない。すなわち、属性値は配列表ファイルの中で一度だけ使用されなければならない。	任意

77. 生物に関する qualifier、すなわち塩基配列の場合は「organism」（附属書 I、第 6 節参照）、アミノ酸配列配列の場合は「organism」（附属書 I、第 8 節参照）は、その配列のソース、すなわち単一の生物または由来を開示しなければならない。生物名は、生物分類データベースから選択されるべきである。

78. 配列が自然発生的であり、由来する生物がラテン語の属名と種名を有する場合は、その名称を qualifier の値として使用しなければならない。推奨する英語の俗名は、塩基配列およびアミノ酸配列とともに qualifier 「note」 を用いて明示しても良いが、organism qualifier の値として使用してはならない。

79. 次の例では、上記の第 77 項および第 78 項に従った配列の由来生物を示す。

#### 例 1：塩基配列の由来

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..5164</INSDFeature_location>
    <INSDFeature_quals>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Solanum lycopersicum</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>note</INSDQualifier_name>
        <INSDQualifier_value>common name: tomato</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>genomic DNA</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_quals>
  </INSDFeature>
</INSDSeq_feature-table>
```

#### 例 2：アミノ酸配列の由来

```
<INSDSeq_feature-table>
  <INSDFeature>
```

```

<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..174</INSDFeature_location>
<INSDFeature_quals>
    <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>protein</INSDQualifier_value>
    </INSDQualifier>
</INSDFeature_quals>
</INSDFeature>
</INSDSeq_feature-table>

```

- 8 0. 配列が自然発生的で、その由来となる生物が既知のラテン語属であるが、種が特定されていない、または未確認である場合、その生物の qualifier 値はラテン語属の後に「sp.」を付さなければならない。例えば、以下のようにになる。

```

<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Bacillus sp.</INSDQualifier_value>

```

- 8 1. 配列は自然発生的であるが、生物のラテン語の属名や種名が不明の場合、生物の qualifier 値は「unidentified」と表示しなければならない。既知の分類学上の情報は、塩基配列およびアミノ酸配列とともに qualifier 「note」で示すべきである。例えば、次のようにになる。

```

<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>unidentified</INSDQualifier_value>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>bacterium B8</INSDQualifier_value>

```

- 8 2. 配列は自然発生的であり、ウイルスのように元となる生物がラテン語の属名と種名を有しない場合は別の許容できる学名（例として、「犬伝染性肝炎 2 型」）を生物の qualifier の値として使用しなければならない。例えば、次のようにになる。

```

<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Canine adenovirus type 2</INSDQualifier_value>

```

- 8 3. 配列が自然発生的で無い場合、生物の qualifier 値を「synthetic construct」と表示しなければならない。配列の生成方法に関するさらなる情報は、塩基配列およびアミノ酸配列とともに qualifier 「note」を用いて示しても良い。例えば、次のようにになる。

```

<INSDSeq_feature-table>
    <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..40</INSDFeature_location>
        <INSDFeature_quals>
            <INSDQualifier>
                <INSDQualifier_name>organism</INSDQualifier_name>
                <INSDQualifier_value>synthetic construct</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                <INSDQualifier_name>mol_type</INSDQualifier_name>
                <INSDQualifier_value>protein</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                <INSDQualifier_name>note</INSDQualifier_name>
                <INSDQualifier_value>synthetic peptide used as assay for
antibodies</INSDQualifier_value>
            </INSDQualifier>
        </INSDFeature_quals>
    </INSDFeature>
</INSDSeq_feature-table>

```

```

</INSDFeature_quals>
</INSDFeature>
</INSDSeq_feature-table>

```

8 4. 塩基配列のqualifier「mol\_type」（附属書I、第6節参照）と、アミノ酸配列の qualifier「mol\_type」（附属書I、第8節参照）は、配列で示される分子のタイプを開示しなければならない。これらの qualifierは、第54項で説明した要素 INSDSeq\_moltype とは異なる。

(a) 塩基配列の場合、qualifier「mol\_type」の値は、以下のいずれかでなければならない。  
 「genomic DNA」、「genomic RNA」、「mRNA」、「tRNA」、「rRNA」、「other RNA」、「other DNA」、「transcribed RNA」、「viral cRNA」、「unassigned DNA」、または「unassigned RNA」。配列が自然発生的で無い場合は、すなわちqualifier「organism」の値が「synthetic construct」である場合、qualifier「mol\_type」の値は「other RNA」または「other DNA」のいずれかでなければならない。

(b) アミノ酸配列の場合、qualifier「mol\_type」の値は「protein」である。

## フリーtekスト

8 5. 第3項(n)に記載のとおり、フリーtekストは特定の qualifier（附属書Iの記載のとおり）値のフォーマットの一種であり、説明的なテキストフレーズの形式で提示され、望ましくは英語またはその他の指定されたフォーマット（附属書I記載のとおり）で提示される。

8 6. フリーtekストの使用は、配列の特性を理解するために不可欠ないくつかの短い用語に限定されていなければならない。qualifier「translation」を除く各 qualifierについて、フリーtekストの文字数は1,000文字を超えてはならない。

8 7. 第3項(o)に記載のとおり、言語に依存したフリーtekストは、国際的、国内的、または広域的な手順の翻訳が必要になる可能性があるという点で、言語に依存する特定の qualifierのフリーtekストの値である。言語に依存するフリーtekスト値のフォーマットを持つ塩基配列の qualifierは、附属書Iの第6節、表5に示されている。言語に依存したフリーtekスト形式のアミノ酸配列の qualifierは、附属書Iの第8節、表6に記載されている。

- (a) 言語に依存するフリーtekストは、英語では要素 INSDQualifier\_value で、英語以外の言語では要素 NonEnglishQualifier\_value で、またはその両方の要素で提示しなければならない。なお、生物名がラテン語の属名および種名である場合には、翻訳は必要ないことに留意されたい。国際的に使用されている英語以外の言葉に由来する専門用語や固有名詞は、要素 INSDQualifier\_value の値の目的のために英語とみなされる（例として、「in vitro」、「in vivo」）。
- (b) 要素 NonEnglishQualifier\_value が配列表に存在する場合、ルート要素 nonEnglishFreeTextLanguageCode 属性に適切な言語コード（第9項の「国際標準化機構(ISO)言語コード 639-1:2002 年公表」参照を参照）が示されなければならない（第43項参照）。单一の配列表内のすべての要素NonEnglishQualifier\_value は、nonEnglishFreeTextLanguageCode 属性で示された言語の値を持たなければならない。要素 NonEnglishQualifier\_value は、言語に依存したフリーtekストの値のフォーマットを有する qualifier に対してのみ許可される。
- (c) NonEnglishQualifier\_value と INSDQualifier\_value が1つの qualifier に対して共に存在する場合、2つの要素に含まれる情報は同等でなければならない。すなわち、NonEnglishQualifier\_value に INSDQualifier\_value の値の翻訳が含まれている、または、INSDQualifier\_value に NonEnglishQualifier\_value の値の翻訳が含まれている、または、両方の要素に originalFreeTextLanguageCode 属性で指定された言語からの qualifier 値の翻訳が含まれている、という条件のいずれかが真でなければならない（第43項参照）。
- (d) 言語に依存するフリーtekスト値を有する qualifier の場合、要素INSDQualifier はオプションの属性 ID を含んでも良い。この属性の値は、「q」の後に「q23」のような正の整数を続けた形式でなければならないらず、1つの要素 INSDQualifier に固有でなければならない。つまり、属性値は、配列表ファイルで1度のみの使用としなければならない。

8 8. 次の例は、第86項の2で説明された言語依存のフリーtekストの提示を示す。

例 1：要素 INSDQualifier\_value における言語依存のフリーtekスト

```
<INSDFeature>
  <INSDFeature_key>regulatory</INSDFeature_key>
  <INSDFeature_location>1..60</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q1">
      <INSDQualifier_name>function</INSDQualifier_name>
      <INSDQualifier_value>binds to regulatory protein Est3</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例 2：要素 INSDQualifier\_value と要素NonEnglishQualifier\_value における言語依存のフリー テキスト

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q45">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>cleaves carbohydrate chain</INSDQualifier_value>
    <NonEnglishQualifier_value>clive la chaîne glucidique</NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例 3：要素 NonEnglishQualifier\_value における言語依存のフリー テキスト。

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q1034">
      <INSDQualifier_name>note</INSDQualifier_name>
      <NonEnglishQualifier_value>clive la chaîne glucidique </NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

## コード配列

8 9. feature key 「CDS」はコード配列、すなわち蛋白質のアミノ酸の配列に対応する塩基配列と終止コドンを識別するために使用しても良い。必須要素 INSDFeature\_location における feature 「CDS」 location には、終止コドンが含まれていなければならない。

9 0. qualifier 「transl\_table」および「translation」は、feature key 「CDS」と共に用いても良い（附属書 I 参照）。qualifier 「transl\_table」を使用しない場合は、標準コード表（附属書 I、第 9 節、表 5 参照）の使用が想定される。

9 1. qualifier 「transl\_except」は、ピロリジン、またはセレノシスティンのいずれかをコード化するコドンを識別するために、feature key 「CDS」および qualifier 「translation」と共に使用しなければならない。

9 2. 第 7 項に包含される qualifier 「translation」に開示され、コード配列によってコードされたアミノ酸配列は、配列表に記載され、個別の配列番号が割り当てられなければならない。アミノ酸配列に割り当てられた配列番号は、feature key 「CDS」を有する qualifier 「protein\_id」の値として記載されなければならない。アミノ酸配列に用いる feature key 「source」の qualifier 「organism」は、そのコード配列のものと同一でなければならない。例えば、次のように。

```
<INSDFeature>
  <INSDFeature_key>CDS</INSDFeature_key>
```

```

<INSDFeature_location>1..507</INSDFeature_location>
<INSDFeature_quals>
  <INSDQualifier>
    <INSDQualifier_name>transl_table</INSDQualifier_name>
    <INSDQualifier_value>11</INSDQualifier_value>
  </INSDQualifier>
  <INSDQualifier>
    <INSDQualifier_name>translation</INSDQualifier_name>
    <INSDQualifier_value>MLVHLERTTIMFDSSLINLPLIWGLLIAIAVLLYILMDGFDLGI
    GILLPFAPSDKCRDHMISSIAPFDGNETWLVLGGGLFAAFPLAYSILMPAFYIPIIM
    LLGLIVRGVSFEFRFKAEGKYRRLWDYAFHFGSLGAAFCQGMILGAFIHGVENGRN
    FSGGQLM
    </INSDQualifier_value>
  </INSDQualifier>
  <INSDQualifier>
    <INSDQualifier_name>protein_id</INSDQualifier_name>
    <INSDQualifier_value>89</INSDQualifier_value>
  </INSDQualifier>
</INSDFeature_quals>
</INSDFeature>

```

### 変異体

- 9 3. 一次配列とその変異体配列、それぞれがそれらの残基の数え上げにより開示され、第 7 項に含まれるのは、それぞれ配列表に含め、個別の配列番号を割り当てなければならない。
- 9 4. 1 つまたは複数の位置に数え上げられた代替残基を有する单一配列として開示された任意の変異体配列は、配列表に記載しなければならず、連続した代替残基が最も限定的な曖昧さ記号で表された单一配列で表示すべきである（第 15 項および第 27 項を参照のこと）。
- 9 5. 一次配列における欠失、挿入または置換によってのみ開示される任意の変異体配列は、配列表に記載すべきである。配列表に含まれる場合、そのような変異体配列は、
- 单一の location または複数の異なる location に変異があり、それらの変異の発生が独立している場合、一次配列の注釈を付して表しても良い。
  - 複数の異なる location に変異があり、それらの変異の発生が相互に依存している場合には、別の配列として表現し、個別の配列番号を割り当てるべきである。
  - 1,000 を超える残基を含む挿入または置換された配列を含む場合は、別の配列として表現し、個別の配列番号を割り振るべきである（第 86 項を参照のこと）。

- 9 6. 次の表は、核酸およびアミノ酸配列の変異体の feature key と qualifier の適切な使用方法を示している。

配列の種類	feature key	qualifier	利用
核酸	variation	replace または note	自然発生の変異および多型、例. 対立遺伝子、制限酵素セグメント長多型(RFLP)
核酸	misc_difference	replace または note	人工的に導入された多様性、例. 遺伝子操作または化学合成
アミノ酸	VAR_SEQ	note	選択的スプライシング、プロモーターの選択的利用、新規転写の開始およびリボソームフレームシフトにより生じる変異体
アミノ酸	VARIANT	note	「VAR_SEQ」が使用できない任意の変異体の種類

- 9 7. 上の表に示されているように、特定の変異体に対する配列の注釈には、feature key、qualifier および

および feature location を含まなければならぬ。qualifier 「replace」の値は、第 1 節、表 1 に記載された記号のみを使用した单一の代替ヌクレオチドまたは塩基配列のみでなければならぬ、または空でなければならぬ。qualifier 「note」の値として、代替残基のリストを記載しても良い。特に、代替アミノ酸のリストは、qualifier 「note」の値として記載しなければならぬ。ここで、配列に「X」が用いられており、「X」は「A」、「R」、「N」、「D」、「C」、「Q」、「E」、「G」、「H」、「I」、「L」、「K」、「M」、「F」、「P」、「O」、「S」、「U」、「T」、「W」、「Y」、または「V」のいずれか 1 つ」以外の値を示す(第 27 項参照)。欠失している場合は、qualifier 「replace」の値を空白にするか、qualifier 「note」にその残基が欠失している可能性があることを示す表示をしなければならぬ。挿入または置換された残基は、qualifier 「replace」または qualifier 「note」で提示されなければならない。qualifier 「replace」および「note」の値の形式は、第 86 項で規定されているように、フリーテキストであり、1,000 文字を超えてはならない。qualifier の値に挿入または置換として記載されている第 7 項に記載されている配列については、第 100 項を参照のこと。

9 8. 附属書 I に記載されている記号（第 1 節から 4 節、表 1 から 4 をそれぞれ参照）は、必要に応じて変異体の残基を表示するために使用されるべきである。qualifier 「note」に関しては、変異体の残基が附属書 I の表 2 または表 4 に記載されていない修飾残基である場合、qualifier 値として修飾残基の省略されていない完全な名称を記載しなければならぬ。修飾残基は、第 17 項または第 30 項に記載があるように、feature table に詳細に記載しなければならぬ。

9 9. 次の例では、上記の第 95 項から 98 項に準拠型の変異体の表示を示す。

例 1：数え上げられた代替ヌクレオチドに用いる feature key 「misc\_difference」。  
配列の 53 位の「n」は、5 つの代替ヌクレオチドのうちの 1 つである。

```
<INSDFeature>
<INSDFeature_key>misc_difference</INSDFeature_key>
<INSDFeature_location>53</INSDFeature_location>
<INSDFeature_quals>
  <INSDQualifier>
    <INSDQualifier_name>note</INSDQualifier_name>
    <INSDQualifier_value>w, cmnm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
  </INSDQualifier>
</INSDFeature_quals>
</INSDFeature>
<INSDFeature>
<INSDFeature_key>modified_base</INSDFeature_key>
<INSDFeature_location>53</INSDFeature_location>
<INSDFeature_quals>
  <INSDQualifier>
    <INSDQualifier_name>mod_base</INSDQualifier_name>
    <INSDQualifier_value>OTHER</INSDQualifier_value>
  </INSDQualifier>
  <INSDQualifier>
    <INSDQualifier_name>note</INSDQualifier_name>
    <INSDQualifier_value>cmnm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
  </INSDQualifier>
</INSDFeature_quals>
</INSDFeature>
```

例 2：塩基配列における欠失に用いる feature key 「misc\_difference」。  
配列の 413 位のヌクレオチドが欠失している。

```
<INSDFeature>
<INSDFeature_key>misc_difference</INSDFeature_key>
<INSDFeature_location>413</INSDFeature_location>
<INSDFeature_quals>
  <INSDQualifier>
    <INSDQualifier_name>replace</INSDQualifier_name>
    <INSDQualifier_value></INSDQualifier_value>
  </INSDQualifier>
```

```
</INSDFeature_quals>
</INSDFeature>
```

例3：塩基配列における挿入に用いる feature key 「misc\_difference」。  
一次配列の 100 位と 101 位の間に、配列「atgccaaatat」が挿入されている。

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>100^101</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>atgccaaatat</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例4：塩基配列における置換に用いる feature key 「variation」。  
シトシンが配列の 413 位のヌクレオチドに置換している。

```
<INSDFeature>
  <INSDFeature_key>variation</INSDFeature_key>
  <INSDFeature_location>413</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>c</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例5：アミノ酸配列において置換に用いる feature key 「VARIANT」。  
配列の 100 位のロケーションにあるアミノ酸は、I, A, F, Y, alle, Melle またはNle で置換することができる。

```
<INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>I, A, F, Y, alle, Melle, or Nle
    </INSDQualifier_value>
  </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>alle, Melle, or Nle</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

例6：アミノ酸配列において置換に用いる feature key 「VARIANT」。  
配列の 100 位に指定されたアミノ酸は、Lys, Arg またはHis を除くすべてのアミノ酸で置換することができる。

```
<INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>not K, R, or H</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

100. 第 7 項に含まれる配列で、一次配列の注釈の qualifier 値の挿入または置換として記載される  
ものも配列表に記載され、個別の配列番号が割り当てられなければならない。

[次頁から附属書 I が続く]

附属書 I

統制語彙

*Version 1.7*

2023年12月8日開催の第11回WIPO標準委員会(CWS)会合で採択

目 次

- 第1節：ヌクレオチドのリスト
- 第2節：修飾塩基のリスト
- 第3節：アミノ酸のリスト
- 第4節：修飾アミノ酸のリスト
- 第5節：塩基配列の feature keys
- 第6節：塩基配列の qualifiers
- 第7節：アミノ酸配列の feature keys
- 第8節：アミノ酸配列の qualifiers
- 第9節：遺伝コード表

## 第 1 節：ヌクレオチドのリスト

配列表で使用されるヌクレオチド塩基の記号を表 1 に示す。記号「t」は、詳細な説明なしに使用される場合、DNA ではチミン、RNA ではウラシルと解釈される。曖昧さ記号（2つ以上の塩基を表す）が適切である場合は、最も限定的な記号を用いるべきである。例えば、ある位置の塩基が「a または g」である可能性がある場合、「n」ではなく「r」を使用すべきである。記号「n」は、それ以上の記述がない状態で使用された場合、「a または c または g または t/u」と解釈される。

表 1：ヌクレオチド記号のリスト

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

## 第 2 節：修飾塩基のリスト

表 2 で示されている略語は、qualifier 「mod\_base」 に許される唯一の値である。特定の修飾されたヌクレオチドが以下の表に存在しない場合は、その値として「OTHER」という略語を使用しなければならない。省略形が「OTHER」の場合は、修飾塩基の省略されていない完全な名称を qualifier 「note」 で指定しなければならない。表 2 に示されている略語は、配列自体に使用してはならない。

表 2：修飾塩基のリスト

Abbreviation	Definition
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
dhu	dihydouridine
fm	2'-O-methylpseudouridine
gal q	beta-D-galactosylqueuosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m4c	N4-methylcytosine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine
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

<b>Abbreviation</b>	<b>Definition</b>
s4u	4-thiouridine
m5u	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u
OTHER	(requires note qualifier)

### 第3節：アミノ酸のリスト

配列に使用するアミノ酸記号を表3に示す。曖昧さ記号（2つ以上のアミノ酸を表す）が適切な場合は、最も限定的な記号を使用すべきである。例えば、ある位置のアミノ酸がアスパラギン酸またはアスパラギンである可能性がある場合には、「X」ではなく「B」という記号を使用すべきである。記号「X」は、それがさらなる説明なしに使用される場合、「A」、「R」、「N」、「D」、「C」、「Q」、「E」、「G」、「H」、「I」、「L」、「K」、「M」、「F」、「P」、「O」、「S」、「U」、「T」、「W」、「Y」、または「V」のいずれか1つと解釈される。

表3：アミノ酸記号のリスト

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

#### 第 4 節：修飾アミノ酸のリスト

表 4 は、feature key 「MOD\_RES」 または「SITE」 の必須 qualifier 「note」において、修飾アミノ酸に許される唯一の略語の一覧である。qualifier 「note」 の値は、必要に応じてこの表の略語か、修飾アミノ酸の省略されていない完全な名称のいずれかでなければならない。この表に記載されている略語(または完全な名称)は、配列自体に使用してはならない。

表 4：修飾アミノ酸のリスト

<b>Abbreviation</b>	<b>Modified Amino acid</b>
Aad	2-Amino adipic acid
bAad	3-Amino adipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
lde	Isodesmosine
alle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
Melle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

## 第 5 節：塩基配列のfeature keys

本節には、塩基配列に使用可能な feature key のリストおよび、その必須・任意の qualifier を示す。feature key はアルファベット順に並んでいる。feature key は、「分子範囲」で指定されていない限り、DNA または RNA に使用することができる。特定の feature key は、「生物範囲」で指定されている意外にも、人工的な配列に使用するのに適している場合がある。

「feature key」の名称は、3'非翻訳領域および 5'非翻訳領域の feature key を除き、以下、記述の「feature key」の後に表示されている通り、配列表の XML インスタンスで使用しなければならない。3'非翻訳領域および 5'非翻訳領域の feature key については、説明文の「コメント」を参照のこと。

5.1	feature key 定義	C_region 免疫グロブリンの軽鎖、重鎖、および t-細胞受容体の $\alpha$ 鎖、 $\beta$ 鎖、 $\gamma$ 鎖の定常領域で、特定の鎖に応じて 1 つまたそれ以上のエクソンを含む。
	任意の qualifier	Allele Gene gene_synonyms Map Note Product pseudo pseudogene standard_name 真核生物
5.2.	feature key 定義	CDS コード配列；蛋白質のアミノ酸配列に対応する塩基配列 (feature location は終止コドンを含む)；feature はアミノ酸の概念的な翻訳が含むことがある。
	任意の qualifier	allele circular_RNA codon_start EC_number exception function gene gene_synonym map note number operaon product Protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_except transl_table
	コメント	qualifier 「codon_start」は、1、2、3 のいずれかの有効な値を持ち、ある code feature の最初の塩基に呼応して、その feature の最初の完全なコドンが確認できるオフセットを示している；qualifier 「transl_table」は、標準的またはユニバーサルな遺伝コード表以外の場合に使用される遺伝コード表を定義する。qualifier 「transl_except」には指定された表の範囲外の遺伝子コードが記載されている。feature key 「CDS」と共に用いることが出来るのは、translation qualifier、「pseudogene」、「pseudo」のいずれか 1 つである。translation qualifier が用いられるとき、翻訳産物に 4 つ以上の具体的に定義されたアミノ酸が含まれる場合、qualifier 「protein_id」が必須となる。
5.3.	feature key	centromere

	定義	セントロメアと同定された生物学的な関心領域で、実験的に特徴づけられているもの。
	任意の qualifier	note standard_name
	コメント	feature 「centromere」は、染色体が保持され、動原体が形成される領域に対応するDNAの区間を示すものである。
5.4.	feature key	D-loop
	定義	置換ループ：ミトコンドリア DNA 内の二本鎖の一方に相同な配列の一本鎖が対合をなす領域；RecA 蛋白質の触媒反応において、二本鎖 DNA の一方が外来の一本鎖によって置換する部位の記述にも用いられる。
	任意の qualifier	allele gene gene_synonym map note
	分子の分類	DNA
5.5.	feature key	D_segment
	定義	免疫グロブリン重鎖および t-細胞受容体の β 鎖の多様性セグメント。
	任意の qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物の分類	真核生物
5.6.	feature key	exon
	定義	スプライシング後の mRNA、rRNA、tRNA の一部をコード化するゲノムの領域；5'非翻訳領域、すべての CDS、3'非翻訳領域を含む場合がある。
	任意の qualifier	allele EC_number function gene gene_synonym map note number product pseudo pseudogene standard_name trans_splicing
5.7.	feature key	gene
	定義	遺伝子として特定され、名称が付された生物学的な関心領域。
	任意の qualifier	allele function gene gene_synonym map note operon pseudo pseudogene phenotype standard_name trans_splicing

	コメント	feature 「gene」は、遺伝的形質や表現型に対応する DNA の区間を記述する。 feature は、定義上、末端の位置に厳密に拘束されるものではない；遺伝子が位置する領域を表すことを意味する。
5.8.	feature key 定義 任意の qualifier	iDNA DNA の介在配列；数種類の組み換えにより排除された DNA。 allele function gene gene_synonym map note number standard_name  分子の分類 コメント 例、免疫グロブリン遺伝子の体細胞処理において。
5.9.	feature key 定義 任意の qualifier	inttron 転写されるが、その配列（エクソン）の両端のどちらか片側をスプライシングすることにより転写産物から取り除かれる DNA のセグメント。 allele function gene gene_synonym map note number pseudo pseudogene standard_name trans_splicing
5.10.	feature key 定義 任意の qualifier	J_segment 免疫グロブリンの軽鎖と重鎖、および t-細胞受容体の $\alpha$ 、 $\beta$ および $\gamma$ 鎖の結合部。 allele gene gene_synonym map note product pseudo pseudogene standard_name
5.11.	生物の分類 feature key 定義 任意の qualifier	真核生物 mat_peptide 成熟ペプチドまたは蛋白質のコード配列；翻訳後修飾を受けて成熟した、または最終的なペプチドまたは蛋白質産物のコード配列；その存在位置は（対応する CDS とは異なり）終止コドンを含まない。 allele EC_number function gene gene_synonym map note product pseudo pseudogene standard_name
5.12.	feature key 定義	misc_binding 他の結合 key（「primer_bind」または「protein_bind」）では表現できない別の部位を共有結合または非共有結合する核酸中の部位。

	必須の qualifier 任意の qualifier	bound_moiety allele function gene gene_synonym map note
	コメント	リボソームの結合部位を記載するためには、feature key 「regulatory」およびその qualifier 「regulatory_class」と共に、qualifier 値「ribosome_binding_site」を用いなければならない。
5.13.	feature key 定義	misc_difference 提示された配列とこの位置で示された配列が異なり、他の Difference キー（「variation」「modified_base」）では表現できない。
	任意の qualifier	allele clone compare gene gene_synonym map note phenotype replace standard_name
	コメント	Feature key 「misc_difference」は人為的に導入された変異性（例. 遺伝子操作または化学合成）に関する記載を行うために用いなければならず、欠失、挿入または置換に関する注釈を付すために qualifier 「replace」を用いる。feature key 「variation」は自然発生的な遺伝的変異を記述するために用いなければならない。
5.14.	feature key 定義	misc_feature 他の feature key では表現することができない生物学的な関心領域；新規または希少な feature。
	任意の qualifier	allele function gene gene_synonym map note number phenotype product pseudo pseudogene standard_name
	コメント	この key は、単にある領域にコメントするため、または、他の feature location で使用するために用いるべきではない。
5.15.	feature key 定義	misc_recomb 他の組換え key や source key の qualifier 「proviral」では説明できない二重鎖 DNA の切断と再結合が行われる、一般的、部位特異的、または複製的な組み換えイベントの部位（プロウイルスの）。
	任意の qualifier	allele gene gene_synonym map note recombination_class standard_name
	分子の分類	DNA
5.16.	feature key 定義	misc_RNA 他の RNA キーでは定義できない転写産物や RNA 産物（prim_transcript, precursor_RNA、mRNA、5'非翻訳領域、3'非翻訳領域、exon、CDS、

		sig_peptide、transit_peptide、mat_peptide、intron、polyA_site、ncRNA、rRNA、tRNA)。
	任意の qualifier	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.17.	feature key	misc_structure
	定義	および他の構造(Structure)キー (stem_loop、D-loop) では記述できない二次、三次のヌクレオチド構造や立体配座。
	任意の qualifier	allele function gene gene_synonym map note standard_name
5.18.	feature key	mobile_element
	定義	可動遺伝子因子を含む、ゲノムの領域。
	必須の qualifier	mobile_element_type
	任意の qualifier	allele function gene gene_synonym map note rpt_family rpt_type standard_name
5.19.	featuue key	modified_base
	定義	示されたヌクレオチドは修飾塩基で、示された分子で置換される (qualifier 「mod_base」の値として記載)。
	必須の qualifier	mod_base
	任意の qualifier	allele frequency gene gene_synonym map note
	コメント	必須 qualifier 「mod_base」 の値は、本附属書 I 、第 2 節の修飾塩基の略語として示されている統制語彙のみを用いる。
5.20.	feature key	mRNA
	定義	メッセンジャーRNA ; 5' 非翻訳領域、コード配列(CDS, エクソン)および、3' 非翻訳領域を含む。
	任意の qualifier	allele circular_RNA function gene gene_synonym map note operon product

		pseudo pseudogene standard_name trans_splicing
5.21.	feature key 定義 必須の qualifier 任意の qualifier	ncRNA rRNA および tRNA 以外の非蛋白質コーディング遺伝子で、その機能分子が RNA 転写物であるもの。 ncRNA_class allele function gnene gene_synonym map note operaon product pseudo pseudogene standard_name trans_splicing
	コメント	feature 「ncRNA」は rRNA および tRNA の注釈に用いてはならない。feature key 「rRNA」および feature key 「tRNA」をそれぞれ用いなければならない。
5.22.	feature key 定義 任意の qualifier	N_region 再配列された免疫グロブリンセグメントの間に挿入された余分なヌクレオチド。 allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物の分類	真核生物
5.23.	feature key 定義 必須の qualifier 任意の qualifier	operon 同じ制御配列/プロモーターの制御下にあり、同じ生物学的経路にある遺伝子群を含む多истем性転写物を含む領域。 operon allele function map note phenotype pseudo pseudogene standard_name
5.24.	feature key 定義 任意の qualifier	oriT 転写の起点；接合または可動化の過程で転写が開始される DNA 分子の領域。 allele bound_moiety direction gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq standard_name

	分子の分類 コメント	DNA 「rep_origin」は転写の起点を記載するために用いなければならない；その qualifier 「direction」の値として用いるのが許容される値には left、right および both があるが、feature 「oriT」と共に用いる場合は left または right のみが有効である。；転写の起点は染色体の中に存在する場合がある。；プラスミドには複数の転写の起点が存在する場合がある。
5.25.	feature key 定義 任意の qualifier	polyA_site 転写後ポリアデニル化によりアデニン残基が付加される、RNA 転写物上の部位。 allele gene gene_synonym map note 生物の分類 真核生物および真核生物ウイルス
5.26.	feature key 定義 任意の qualifier	precursor_RNA 未成熟な RNA 産物ではない RNA 種；ncRNA、rRNA、tRNA、5'非翻訳領域、コード配列(CDS、エキソン)、介在配列(イントロン)および、3'非翻訳領域を含んでいても良い。 allele function gene gene_synonym map note operon product standard_name trans_splicing
	コメント	転写後のプロセシングの結果であると思われる RNA に使用される；問題の RNA が、プロセシングされていないのが知られている場合は、key 「prim_transcript」を用いる。
5.27.	feature key 定義 任意の qualifier	prim_transcript (最初の、未転写の)一次転写物；ncRNA、rRNA、tRNA、5' 非翻訳領域、コード配列(CDS、エキソン)、介在配列(イントロン)および、3' 非翻訳領域を含んでいても良い。 allele function gene gene_synonym <u>map</u> <u>note</u> operon standard_name
5.28.	feature key 定義 任意の qualifier	primer_bind 複製、転写または逆転写を開始するための非共有結合型プライマーの結合部位；合成用(例、PCR プライマー要素)の部位を含む。 allele gene gene_synonym map note standard_name コメント プライマー分子が結合する、配列上のある部位の注釈情報の記載に用いる feature key - プライマー分子そのものの配列を示すためには用いない；ポリメラーゼ連鎖(PCR)反応ではプライマー対を使用することが多いため、1つの key 「primer_bind」で 2 つの存在位置を指定して order(存在位置、存在位置)演算子を用いたり、key 「primer_bind」を 2 つ用いても良い。
5.29.	feature key 定義	propeptide プロペプチドコード配列；プロタンパク質のドメインのコード配列で、切断されて成熟蛋白質産物を形成するもの。

	任意の qualifier	allele function gene gene_synonym map note product pseudo pseudogene standard_name
5.30.	feature key 定義 必須の qualifier 任意の qualifier	protein_bind 核酸上の、非共有の蛋白質結合部位。 bound_moiety allele function gene gene_synonym map note operon standard_name
	コメント	リボソームの結合部位を記述するには、feature key 「regulatory」 およびその qualifier 「regulatory_class」 (qualifier 値は「ribosome_binding_site」)を用いなければならない。
5.31.	feature key 定義 必須の qualifier 任意の qualifier	regulatory 転写、翻訳、複製またはクロマチン構造の制御に機能する配列の任意の領域。 regulatory_class allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name
5.32.	feature key 定義 任意の qualifier	repeat_region 繰り返し単位を含むゲノム領域。 allele function gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq satellite standard_name
5.33.	feature key 定義 任意の qualifier	rep_origin 複製の起点；核酸の 2 つの同じ複製を作成するための複製の起点。 allele direction function gene gene_synonym

		map note standard_name qualifier 「direction」 は 3 つの有効値を用いる。left、right または both。
5.34.	コメント feature key 定義	rRNA 成熟したリボソーム RNA；アミノ酸を蛋白質に合成するリボ核蛋白質粒子（リボソーム）を構成する RNA 成分。
	任意の qualifier	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name
	コメント feature key 定義	qualifier 「product」 を用いて rRNA の大きさに関する注釈を付すべきである。
5. 35.	任意の qualifier	S_region 免疫グロブリンの重鎖のスイッチ領域；同じ B 細胞から異なる免疫グロブリンクラスを発現させる重鎖 DNA の再編成に関与する。 allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物の分類	真核生物
5.36.	feature key 定義	sig_peptide シグナルペプチドのコード配列；分泌蛋白質の N 末端ドメインのコード配列；このドメインは、新生ポリペプチドを膜のリーダー配列に付着させることに関与する。 allele function gene gene_synonym map note product pseudo pseudogene standard_name
	任意の qualifier	
5.37.	feature key 定義 必須の qualifier	source この key は必須；すべての配列は、配列全体を網羅する単一のソース(source) key を有する。 organism mol_type cell_line cell_type chromosome clone clone_lib collected_by collection_date cultivar dev_stage
	任意の qualifier	

		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
	分子の分類	すべて
5.38.	feature key 定義	stem_loop ヘアピン構造；RNA または DNA の一本鎖の中で、隣接する（逆）相補的な配列同士が塩基対合して形成された二重螺旋状の領域。
	任意の qualifier	allele function gene gene_synonym map note operon standard_name
5.39.	feature key 定義	STS 配列タグ部；ゲノム状のマッピングランドマークを特徴づける短い、單一コピーの DNA 配列で、PCR によって検出される；一連の配列タグ部(STS)の順序を決定することでゲノムの一領域をマッピングできる。
	任意の qualifier	allele gene gene_synonym map note standard_name
	分子の分類 コメント	DNA key 「primer_bind」 または複数のプライマーに一つまたは複数のプライマーを含めるための配列タグ部(STS)の存在位置。
5.40.	feature key 定義 任意の qualifier	telomere テロメアとして特定され、実験的に特徴づけられた生物学的な関心領域。 note

		rpt_type rpt_unit_range rpt_unit_seq standard_name
	コメント	feature 「telomere」は、真核生物の直線状の染色体の末端にある特定の構造に対応する DNA の区間のことと、末端の完全性と維持に必要とされる；この領域は、他の染色体と比べて特異であり、染色体の物理的な末端を示している。
5.41.	feature key 定義	tmRNA 転移伝令 RNA (tmRNA) ; tmRNA は、まず tRNA として働き、次にペプチドタグをコードする mRNA として働く；リボソームが tmRNA の mRNA 領域を翻訳し、コードされたペプチドタグを未成熟蛋白質の C 末端に結合することで、このタグが蛋白質の破壊や蛋白質分解の標的となる。
	任意の qualifier	allele function gene gene_synonym map note product pseudo pseudogene standard_name tag_peptide
5.42.	feature key 定義	transit_peptide ペプチドのコード配列：核にコードされたオルガネラ蛋白質の N 末端のドメインのコード配列；このドメインは、オルガネラへの蛋白質の翻訳後移入に関与する
	任意の qualifier	allele function gene gene_synonym map note product pseudo pseudogene standard_name
5.43.	feature key 定義	tRNA 成熟した tRNA : 核酸配列からアミノ酸配列への翻訳を媒介する低分子 RNA (長さ 75~85 塩基)。
	任意の qualifier	allele circular_RNA anticodon function gene gene_synonym map note <u>operon</u> product pseudo pseudogene standard_name trans_splicing
5.44.	feature key 定義	unsure 配列された塩基のうち、通常 10 個以下の長さの小さな領域で、明確に特定できないもの。このような領域には、呼び出された塩基 (a, t, g, c) や、呼び出された塩基と呼び出されていない塩基 ('n') が混在していることがある。
	任意の qualifier	allele compare gene

		gene_synonym map note replace
	コメント	欠失、挿入または置換に関する情報を記載する場合は、qualifier 「replace」を用いて注釈を付す。
5.45.	feature key 定義	V_region 免疫グロブリンの軽鎖、重鎖、T細胞受容体の $\alpha$ 、 $\beta$ 、 $\gamma$ 鎖の可変領域で、可変アミノ末端部分をコードし、V_segments、D_segments、N_region、J_segmentsで構成されることが可能。
	任意の qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物の分類	真核生物
5.46.	feature key 定義	V_segment 免疫グロブリンの軽鎖、重鎖、T細胞受容体の $\alpha$ 鎖、 $\beta$ 鎖、 $\gamma$ 鎖の可変セグメントで、可変領域 (V_region) の大部分とリーダーペプチドの末端の数個のアミノ酸をコードする。
	任意の qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物の分類	真核生物
5.47.	feature key 定義	variation 近親株に、同じ遺伝子からの不可逆の突然変異体（制限酵素セグメント長多型 (RFLP)、多型など）があり、この存在位置（および他の存在位置）で提示された配列と異なる場合。
	任意の qualifier	allele compare frequency gene gene_synonym map note phenotype product replace standard_name
	コメント	対立遺伝子（アレル）、制限酵素セグメント長多型(RFLP)、および他の自然発生の突然変異体および多型について記載するために用いる。欠失、挿入、置換を示す注釈を付すためには qualifier 「replace」を用いる；遺伝子操作（例、部位具体的変異誘発(SDM)）の結果生じる変異性は、feature key 「misc_difference」を用いて記載しなければならない。
5.48.	feature key 定義	3'UTR 1) 蛋白質に翻訳されない成熟した転写物の 3'末端（終止コドンの後）の領域。 2) RNA ウィルスの 3' 末端の領域（最後の終止コドンの後）で、蛋白質に翻訳されない領域。
	任意の qualifier	allele function

---

	gene gene_synonym map note standard_name trans_splicing	
コメント		アポストロフィー文字(「'」)は、XML 形式の文書では特別の意味があり、要素の値として記載する際は「&apos;」で置き換えなければならない。したがって、「3'非翻訳領域」は XML ファイルでは「3&apos;UTR」、つまり<INSDFeature_key>3&apos;UTR</INSDFeature_key>と記載しなければならない。
5.49. feature key	5'UTR	
定義		1) タンパク質に翻訳されない成熟転写産物の 5'末端（開始コドンの前）の領域。 2) タンパク質に翻訳されない RNA ウィルスの 5'末端の領域（第 1 開始コドンの前）。
任意の qualifier	allele function gene gene_synonym map note standard_name trans_splicing	
コメント		アポストロフィー文字(「'」)は、XML 形式の文書では特別の意味があり、要素の値として記載する際は「&apos;」と置き換えなければならない。したがって、「5'非翻訳領域」は XML ファイルでは「5&apos;UTR」、つまり<INSDFeature_key>5&apos;UTR</INSDFeature_key>と記載しなければならない。

## 第 6 節 塩基配列の qualifiers

本節では、塩基配列のfeatureに使用される qualifier のリストを掲載している。qualifierはアルファベット順に並んでいる。

qualifier 値のフォーマットが「none」の場合、要素「INSDQualifier\_value」は使用してはならず、要素 NonEnglishQualifier\_value も使用してはならない。

値のフォーマットが言語に依存するフリーテキストの場合、以下のいずれかを使用しなければならない。

- 1) 要素 INSDQualifier\_value、または
- 2) 要素 NonEnglishQualifier\_value、または
- 3) 要素 INSDQualifier\_value と要素 NonEnglishQualifier\_value の両方。

値のフォーマットが「none」以外で、言語依存のフリーテキストとして識別されない場合、要素 INSDQualifier\_value を使用しなければならず、要素 NonEnglishQualifier\_value は使用してはならない。

注意：言語に依存するフリーテキスト値形式を有する qualifier に対して提供される qualifier 値は、国際的、国内的、または広域的な手続きのために翻訳が必要になる場合がある。次の表に示す qualifier は、言語に依存したフリーテキストの値を持つとみなされる。

表 5：言語に依存するフリーテキスト値を持つ塩基配列の qualifier のリスト

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

6.1.	qualifier	allele
	定義	特定の遺伝子の対立遺伝子に付した名称
	必須 qualifier 値のフォーマット	フリーテキスト
	例	<INSDQualifier_value>adh1-1</INSDQualifier_value>
	コメント	ある遺伝子に存在する、遺伝子関連の feature (エクソン、コード領域等) にはすべて、同じ qualifier 「allele」 の qualifier 値を用いるべきである；qualifier 「allele」 の qualifier 値は、定義上 qualifier 「gene」 の qualifier の値と同じであってはならない；feature key 「variation」 と共に用いられる場合、qualifier 「allele」 の qualifier 値は、その変異体の値であるべきである。
6.2.	qualifier	anticodon
	定義	tRNA のアンチコドンの位置とそれがコードするアミノ酸の位置
	必須 qualifier 値のフォーマット	pos:<location>,aa:<amino_acid>,seq:<text> <location>は、アンチコドンの位置 <amino_acid>は、tRNA がコードするアミノ酸の 3 文字の略語 <text>は、アンチコドンの配列
	例	<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
	定義	特定の feature と結合する可能性のある分子/複合体の名称
	必須 qualifier 値のフォーマット	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	例	<INSDQualifier_value>GAL4</INSDQualifier_value>
	コメント	「misc_binding」、「riT」、「protein_bind」の各 feature には、1 つの qualifier 「bound_moiety」 の使用が可能。
6.4.	qualifier	cell_line
	定義	配列が得られた細胞株
	必須 qualifier 値のフォーマット	フリーテキスト
	例	<INSDQualifier_value>MCF7</INSDQualifier_value>
6.5.	qualifier	cell_type
	定義	配列が得られた細胞の種類
	必須 qualifier 値のフォーマット	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	例	<INSDQualifier_value>leukocyte</INSDQualifier_value>
6.6.	qualifier	chromosome
	定義	配列が得られた染色体 (例. 染色体番号)
	必須 qualifier 値のフォーマット	フリーテキスト
	例	<INSDQualifier_value>1</INSDQualifier_value> <INSDQualifier_value>X</INSDQualifier_value>
6.7	qualifier	circular_RNA
	定義	例えば、遺伝子の下流のエクソンが RNA 産物の上流のエクソンの 5'に位置している場合など、このスプライスされた RNA 産物がバックスプライシングによって生じた環状 RNA (circular RNA) であるため、エクソンの順番がずれていったり、重なっていたりすることを示している。
	qualifier 値のフォーマット	なしあり
	コメント	CDS, mRNA, tRNA など、バックスプライシングの結果として生成される feature に使用する必要がある。この qualifier は、スプライスイベントが演算子「join」で示されている場合にのみ使用する必要がある (例 : join(complement(69611..69724),139856..140087))
6.8.	qualifier	clone および

	定義 必須 qualifier 値の フォーマット 例 コメント	配列が得られたクローン フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value> ソース(source) feature には、複数の qualifier 「clone」 を用いてはならない；複数のクローンから得られた配列の場合は、feature key 「misc_feature」 およびその qualifier 「note」 を用いて feature table に記載しても良い。
6.9.	qualifier 定義 必須 qualifier 値の フォーマット 例	clone_lib 配列が得られたクローンライブラリ フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>lambda-hIL7</INSDQualifier_value>
6.10.	qualifier 定義 必須 qualifier 値の フォーマット 例	codon_start その feature の最初の塩基に呼応して、コーディングされた feature の最初の完全なコドンが確認できるオフセットを示している。 1 または 2 または 3 <INSDQualifier_value>2</INSDQualifier_value>
6.11.	qualifier 定義 必須 qualifier 値の フォーマット 例	collected_by 標本個体を採取した人の名前や機関の名称 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Dan Janzen</INSDQualifier_value>
6.12.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	collection_date 標本個体が採取された日付 YYYY-MM-DD、YYYY-MM または、YYYY のみ <INSDQualifier_value>1952-10-21</INSDQualifier_value> <INSDQualifier_value>1952-10</INSDQualifier_value> <INSDQualifier_value>1952</INSDQualifier_value> 「YYYY」は暦年を表す 4 桁の数字、「MM」は暦月を表す 2 桁の数字、「DD」はある暦月の暦日を示す 2 桁の数字
6.13.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	compare 比較対象となる、公開されている国際塩基配列データベースの参照情報 [accession-number.sequence-version] <INSDQualifier_value>AJ634337.1</INSDQualifier_value> この qualifier は、「misc_difference」「unsure」および「variation」の各 feature で使用しても良い。1 つのフィーチャーに、内容の異なる複数の qualifier 「compare」 の使用が許されている。この qualifier は、一塩基多型(SNP)のような大規模な変異体に注釈を付すことを想定していない。
6.14.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	cultivar 配列が得られた植物の栽培品種 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Nipponbare</INSDQualifier_value> <INSDQualifier_value>Tenuifolius</INSDQualifier_value> <INSDQualifier_value>Candy Cane</INSDQualifier_value> <INSDQualifier_value>IR36</INSDQualifier_value> qualifier 「cultivar」 は、人為的に選択された産物のみに適用される；自然の、名前の付いた植物や真菌類には qualifier 「variety」 を用いる。
6.15.	qualifier 定義	dev_stage 配列が特定の発達段階にある生物から得られたものであれば、この qualifier で指定される。

	必須 qualifier 値の フォーマット 例	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>fourth instar larva</INSDQualifier_value>
6.16.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	direction DNA複製の方向 left, right または both left とは(配列上で示されている)5'末端側の方向を示し、right とは 3'末端側の方向を示している。 <INSDQualifier_value>left</INSDQualifier_value> feature key 「rep_origin」に qualifier 「direction」を使用した場合、left、right、both の値が許容される。ただし、feature key 「oriT」に qualifier 「direction」を使用した場合は、left および right の値のみが許可される。 rep_origin 機能のキーに direction qualifier を使用した場合、left、right、both の値が許容される。
6.17.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	EC_number 配列の酵素生成物の酵素コミュニケーション番号 フリーテキスト <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> 国際生化学分子生物学連合の命名法委員会(NC-IUBMB)が作成したリスト中で定義されている酵素委員会番号(CE 番号)の有効値 (1992 年発表の酵素命名法勧告、アカデミックプレス、サンディエゴまたは、最新改訂版に掲載)。酵素委員会番号はピリオドで区切られた 4 つ数字の文字列を表し、文字列の最後から 3 つまでの数字は、割り当て不確かであることを示すためにダッシュ「-」で置き換えることができる。「n」を含む記号(例:「n」、「h1」など)は、EC 番号が割り当て待ちである場合に、数字の代わりに最後の位置に使用することができる。このような不完全な EC 番号は NC-IUBMB では承認されないため注意を要する。
6.18.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	ecotype 遺伝学的に生育環境への適応を反映した表現型特性を示す種内集団 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Columbia</INSDQualifier_value> このような集団の例としては、特に日当たりの良い生息環境に対応して、通常の葉よりも毛状突起が多い葉を有する様になった集団が挙げられる。「生態型」(ecotype)は、シロイヌナズナの標準的な遺伝子株に用いられることが多いが、あらゆる無菌生物に適用することができる。
6.19.	qualifier 定義 qualifier 値のフォ ーマット コメント	environmental_sample 大量の環境中の DNA サンプルから直接分子的に分離して得られた配列(ポリメラーゼ連鎖反応(PCR)とそれに続く産物のクローニング、変性剤濃度勾配ゲル電気泳動法(DGGE)または他の汎用手法による)で、その由来生物を確実に特定できないものを指す。環境サンプルには、臨床検体および腸管内容物、および、特定の宿主に関連する可能性のある無名の生物からの他の配列が含まれる。これらには、特定の宿主から確実に回収できる内共生体、容易に識別できるが培養されていない野外サンプルからの生物(例:多くのシアノバクテリア)、病害が認められる植物から確実に回収できるファイトプラズマ(純粋培養は不可能であるが)は含まれない。 なし feature key 「source」と共にのみ用いられる； qualifier 「environmental_sample」を用いる feature key 「source」には、qualifier 「isolation_source」も用いる必要があり、qualifier 「environmental_sample」を用いる feature 「source」には qualifier 「strain」を用いてはならない。
6.20.	qualifier 定義	exception 標準的な生物学の法則を用いて翻訳できないコード領域であることを示している。

	必須 qualifier 値のフォーマット	以下の統制語彙を用いたフレーズの 1 つを使用する。 RNA editing (ゲノム編集) rearrangement required for product (転写産物が求める再編成) annotated by transcript or proteomic data (転写データまたはプロテオミクスに関するデータを用いた詳細情報の記載)
	例	<INSDQualifier_value>RNA editing</INSDQualifier_value> <INSDQualifier_value>rearrangement required for product</INSDQualifier_value>
	コメント	RNA 演習等の生物学的メカニズムを説明するためのみに用いられる ; qualifier 「exception」を用いた CDS の蛋白質翻訳は、対応する概念的な翻訳とは異なる ; qualifier 「transl_except」を用いるのが適切である場合（例. 翻訳が終止コドンで完了する場合）、この qualifier を用いてはならない。
6.21.	qualifier 定義 必須 qualifier 値のフォーマット	Frequency feature の発生頻度 feature を有する個体群の比率を割合で記載するフリーテキスト 言語依存 : この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	例	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.22.	qualifier 定義 必須 qualifier 値のフォーマット	function 配列に起因する機能 フリーテキスト 言語依存 : この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	例	<INSDQualifier_value>essential for recognition of cofactor </INSDQualifier_value>
	コメント	function qualifier は、遺伝子名や製品名では配列に起因する機能を伝えられない場合に使用される。
6.23.	qualifier 定義 必須 qualifier 値のフォーマット	gene 配列領域に対応する遺伝子記号 フリーテキスト
	例	<INSDQualifier_value>ilvE</INSDQualifier_value>
	コメント	qualifier 「gene」は遺伝子記号を記載するために用いる ; 完全な遺伝子名を記載するには qualifier 「standard_name」を用いる。
6.24.	qualifier 定義 必須 qualifier 値のフォーマット	gene_synonym 同義的な、置換された、廃止された、または以前の遺伝子記号 フリーテキスト 言語依存 : この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	例	<INSDQualifier_value>Hox-3.3</INSDQualifier_value> (qualifier 「gene」の qualifier 値が「Hoxc6」である feature。)
	コメント	遺伝子記号の動議とを示すのに役立つ場合に使用される ; qualifier 「gene_synonym」を用いる場合、主要な遺伝子記号は必ず qualifier 「gene」で記載しなければならない。
6.25.	qualifier 定義 qualifier 値のフォーマット	germline 適応免疫反応の一部として体細胞の再編成を受けていない配列 ; 親の生殖細胞から受け継いだ再編成されていない配列である。 なし
	コメント	qualifier 「germline」、配列の由来がガメートまたは生殖細胞であるとの情報を記載するために用いてはならない ; qualifier 「germline」および qualifier 「rearranged」は同じ source feature 内で使用してはならない ; qualifier 「germline」と qualifier 「rearranged」は、適応免疫反応の一部として体細胞の再編成を受ける可能性のある分子にのみ使用しなければならない ; これらは、頸のある脊椎動物の t-細胞受容体(TCR)と免疫グロブリン遺伝子座、および頸のない魚（ヤツメウナギやメクラウナギ）の無関係な可変リンパ球受容体(VLR)遺伝子座である ; qualifier 「germline」と qualifier 「rearranged」は、有頭動物(系統分類 ID = 89593)以外で使用するべきではない。
6.26.	qualifier	haplogroup

	<b>定義</b>	配列多型を共有する類似したハプロタイプのグループ名称。ハプログループは、個体群移動の追跡によく用いられる。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>H*</INSDQualifier_value>
6.27.	<b>qualifier</b>	<b>haplotype</b>
	<b>定義</b>	同一の物理的染色体上で連鎖する対立遺伝子の組み合わせの名称。組換えがない場合、各ハプロタイプは 1 つの単位として遺伝するため、集団における遺伝子流動を追跡するのに利用されても良い。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト <INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>
6.28.	<b>qualifier</b>	<b>host</b>
	<b>定義</b>	配列された分子が得られた生物の自然界（実験室ではなく）における宿主
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Homo sapiens</INSDQualifier_value> <INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value> <INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>
6.29.	<b>qualifier</b>	<b>identified_by</b>
	<b>定義</b>	標本個体の分類学的な同定を行った専門家の名前
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>John Burns</INSDQualifier_value>
6.30.	<b>qualifier</b>	<b>isolate</b>
	<b>定義</b>	配列が得られた個々の分離
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Patient #152</INSDQualifier_value> <INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>
6.31.	<b>qualifier</b>	<b>isolation_source</b>
	<b>定義</b>	配列が得られた生物学的サンプルの物理的、環境的および／あるいは地理学的由来を記述する。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <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>
	<b>コメント</b>	feature key 「source」でのみ使用する；qualifier 「environmental_sample」を用いる feature key 「source」は、qualifier 「isolation_source」も用いるべきである。
6.32.	<b>qualifier</b>	<b>lab_host</b>
	<b>定義</b>	配列された分子が得られた由来生物を増殖させるために実験室で使われた宿主の学名。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <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>
	<b>コメント</b>	宿主生物の二名法による学名が分かっている場合は、それを使用べきである；宿主に関連する追加情報を記載しても良い。
6.33.	<b>qualifier</b>	<b>lat_lon</b>

	定義 必須 qualifier 値の フォーマット 例	標本個体が採取された場所の地理的座標 フリーテキスト – 経緯度の座標の雑形は「d[d.dddd] N S d[dd.dddd] W E」 <INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value> <INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>
6.34.	qualifier 定義 qualifier 値の フォーマット	macronuclear 示されている配列が DNA であり、大核と小核の間で染色体の分化を行う生物からのものである場合、その配列が大核 DNA からのものであることを示すために、使用する。 なし
6.35	qualifier 定義 必須 qualifier 値の フォーマット 例	map ゲノムマップ上の feature の位置 フリーテキスト <INSDQualifier_value>8q12-q13</INSDQualifier_value>
6.36.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	mating_type 配列が得られた生物の交配タイプ；交配タイプは、原核生物および、雌雄同体の配偶子を持たずに減数分裂を行う真核生物に使用される。 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <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>" qualifier 「mating_type」の male と female は原核生物では有効だが、真核細胞では無効である。詳細は、qualifier 「sex」 のコメントを参照のこと。
6.37.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	mobile_element_type 親 feature で記述された可動因子の識別子のタイプおよび名称 <mobile_element_type>[:<mobile_element_name>] 上記<mobile_element_type>には、以下のいずれかを用いる。 transposon retrotransposon integron insertion sequence non-LTR retrotransposon SINE MITE LINE other <INSDQualifier_value>transposon:Tnp9</INSDQualifier_value> qualifier 「mobile_element_type」は、feature key 「mobile_element」 にのみ許される。可動因子は、現在可動的な因子と、過去に可動的であった因子の両方を表すために使用されるべきである。<mobile_element_type>の qualifier 値 「other」 には、<mobile_element_name>が必要である。
6.38.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	mod_base 修飾塩基の略語 本附属書の第 2 節から、修飾塩基の略語を選択する。 <INSDQualifier_value>m5c</INSDQualifier_value> <INSDQualifier_value>OTHER</INSDQualifier_value> 本附属書の第 2 節にない特定の修飾塩基は、qualifier 「mod_base」 の値として「OTHER」を入力し、その値として修飾塩基の完全名称と共に qualifier 「note」 を記載することで注釈を付す。
6.39.	qualifier	mol_type

---

定義	配列の分子タイプ
必須 qualifier 値の フォーマット	以下より 1 つを選択する。 genomic DNA genomic RNA mRNA tRNA rRNA other RNA other DNA transcribed RNA viral cRNA unassigned DNA unassigned RNA
例	<INSDQualifier_value>genomic DNA</INSDQualifier_value> <INSDQualifier_value>other RNA</INSDQualifier_value>
コメント	qualifier 「mol_type」は feature key 「source」に必須である；本 qualifier の qualifier 値「genomic DNA」は、分子が核であることを意味しない（例. 細胞小器官やプラスミド DNA は、「genomic DNA」を使用して記述する必要がある。）；リボソーム RNA 遺伝子は「genomic DNA」を用いなければならない；「rRNA」は、リボソーム RNA 分子自体の配列が決定されている場合にのみ適用しなければならない；「other RNA」および「other DNA」は合成分子にのみ用いなければならない、「unassigned DNA」および「unassignedRNA」は、生体内の分子が不明な場合に適用しなければならない。

---

6.40.	qualifier	ncRNA_class
	定義	parent key 「ncRNA」によって記述されたノンコーディング RNA の分類の構造化された記述。
	必須 qualifier 値の フォーマット	TYPE
		TYPE とは以下の統制語彙または統制フレーズのうちのいずれか 1 つである。 antisense_RNA autocatalytically_spliced_intron circRNA 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
	例	<INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>siRNA</INSDQualifier_value> <INSDQualifier_value>scRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value>
	コメント	qualifier 「ncRNA_class」に用いられていない特定の ncRNA タイプは、qualifier 「ncRNA_class」に「other」と入力し、qualifier 「note」に新規「ncRNA_class」の簡単な説明を記述することで、注釈を付す必要がある。

6.41.	qualifier	note コメントや追加情報 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>A comment about the feature</INSDQualifier_value>
6.42.	qualifier	number 定義 5'末端から 3'末端の方向に存在する遺伝因子（例. エクソンまたはイントロン）の順序を示す数字。 必須 qualifier 値の フォーマット フリーtekスト（空白（スペース）文字は用いない） 例 <INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value> コメント 文字列には、空白文字を含まないデータ値として表される、整数、文字、または整数と文字の組み合わせに限定される。追加の用語は情報は qualifier「standard_name」で記載すべきである；例：値が「2A」である qualifier「number」や、値が「long」である qualifier「standard_name」。
6.43.	qualifier	operon 定義 その feature が属する、単一の転写物に転写された連続した遺伝子群の名称。 必須 qualifier 値の フォーマット フリーtekスト 例 <INSDQualifier_value>lac</INSDQualifier_value>
6.44.	qualifier	organelle 定義 配列が取得された膜結合型細胞内構造の種類。 必須 qualifier 値の フォーマット 以下の統制語彙および統制句の 1 つを本 qualifier 値として用いる。 chromatophore hydrogenosome mitochondrion nucleomorph plastid mitochondrion:kinetoplast plastid:chloroplast plastid:apicoplast plastid:chromoplast plastid:cyanelle plastid:leucoplast plastid:proplastid 例 <INSDQualifier_value>chromatophore</INSDQualifier_value> <INSDQualifier_value>hydrogenosome</INSDQualifier_value> <INSDQualifier_value>mitochondrion</INSDQualifier_value> <INSDQualifier_value>nucleomorph</INSDQualifier_value> <INSDQualifier_value>plastid</INSDQualifier_value> <INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chloroplast</INSDQualifier_value> <INSDQualifier_value>plastid:apicoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chromoplast</INSDQualifier_value> <INSDQualifier_value>plastid:cyanelle</INSDQualifier_value> <INSDQualifier_value>plastid:leucoplast</INSDQualifier_value> <INSDQualifier_value>plastid:proplastid</INSDQualifier_value>
6.45.	qualifier	organism 定義 配列された遺伝物質を記載した生物の学名（判明している場合）、またはその生物が分類されていない場合には利用可能な分類学的情報；または配列が人工的に構築された旨の表示。 必須 qualifier 値の フォーマット フリーtekスト 例 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
6.46.	qualifier	PCR_primers

	<b>定義</b>	配列の増幅に使用された PCR プライマー。単一の qualifier 「PCR_primers」は、単一の PCR 反応に使用されるすべてのプライマーを用いるべきである。単一の PCR 反応に複数のフォワードプライマーまたはリバースプライマーが存在する場合、複数の fwd_name/fwd_seq や rev_name/rev_seq の組が存在することになる。
	<b>必須 qualifier 値の フォーマット 例</b>	[fwd_name: XXX1, ]fwd_seq: xxxxx1,[fwd_name: XXX2, ]fwd_seq: xxxxx2, [rev_name: YYY1, ]rev_seq: yyyy1,[rev_name: YYY2, ]rev_seq: yyyy2<INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value><INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<&lt;i&&gt;gtgtatcttact</INSDQualifier_value><INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value>
	<b>コメント</b>	fwd_seq と rev_seq はともに必須である ; fwd_name と rev_name はともに任意である。両配列は 5'末端から 3'末端の方向に示さなければならない。配列は本附属書の第 1 節に記載されている記号を用いなければならない、修飾塩基を除いて、山かっこ(すなわち「<」または「>」)の中に記号を記載しなければならない。XML 形式では、山かっこ(「<」および「>」)は、XML の予約文字であるため、「&lt;」と「&gt;」に置き換えなければならない。
6.47.	<b>qualifier 定義</b>	phenotype feature によって与えられる表現型（ここで表現型とは、物理的、生化学的、行動的な特性または一連の特性と定義される）。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>erythromycin resistance</INSDQualifier_value>
6.48.	<b>qualifier 定義</b>	plasmid 配列が得られた天然のプラスミドの名称。プラスミドは、qualifier 「chromosome」 または qualifier 「segment」 では表現できない、独立して複製する遺伝学的単位と定義される。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト <INSDQualifier_value>pC589</INSDQualifier_value>
6.49.	<b>qualifier</b>	pop_variant 配列の由来となったサンプルの亜集団または表現型の名称。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>
6.50.	<b>qualifier 定義</b>	product feature に関連する生成物の名称。例、 feature 「mRNA」 の mRNA、 CDS の polypeptide、 mat_peptide の mature peptide など。
	<b>必須 qualifier 値の フォーマット 例</b>	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>trypsinogen</INSDQualifier_value> (when qualifier appears in CDS feature) <INSDQualifier_value>trypsin</INSDQualifier_value> (when qualifier appears in mat_peptide feature) <INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (when qualifier appears in mRNA feature)
6.51.	<b>qualifier 定義</b>	protein_id 蛋白質配列番号（対応する feature key 「CDS」 およびその translation qualifier で特定されるコード配列によってコードされた蛋白質の配列を指定するために、配列表で使用される整数）。
	<b>必須 qualifier 値の フォーマット 例</b>	ゼロより大きい整数 <INSDQualifier_value>89</INSDQualifier_value>

6.52.	qualifier 定義 qualifier 値のフォーマット	proviral 本 qualifier は、他の生物のゲノムに組み込まれたウイルスやファージから得られた配列にフラグを立てるために使用する。 なし
6.53.	qualifier 定義 qualifier 値のフォーマット コメント	pseudo feature key によって特定されている機能を有していたが、その機能を失った feature であることを示す qualifier。 なし qualifier 「pseudo」は、以前は偽遺伝子と表現されていなかったが現在は機能を失った遺伝子であることを示すために用いられるべきである。例として、偽遺伝子化の事象発生以外の理由で翻訳機能を失ったコード領域。他の理由としては、配列決定や構築の誤りなどが考えられる。偽遺伝子に注釈を付すためには、偽遺伝子のタイプを示す qualifier 「pseudogene」を使用しなければならない。
6.54.	qualifier 定義 必須 qualifier 値のフォーマット TYPE 例 コメント	pseudogene この feature が、feature key で指定された要素の偽遺伝子であることを示す。 上記の TYPE は、以下の統制語彙または統制句のうちの 1 つである。 processed unprocessed unitary allelic unknown <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> 本 qualifier 値「TYPE」の定義は以下の通り。：  処理済み(processed) – 偽遺伝子が、mRNA が cDNA に逆転写された後、ゲノムに再統合されて生じたもの。そのため、イントロンやエクソンの構造は失われており、疑似ポリ(A)尾部を有している場合がある。  未処理(unprocessed) – 偽遺伝子が、親遺伝子のコピーから、複製とその後の変則的な突然変異の蓄積によって生じたもの。機能的の相同意と比較して、挿入、欠失、未成熟終止コドン、フレームシフトおよび同義置換の逆である非同義置換の割合増加等の変化が見られる。  単一(unitary) – 親遺伝子が存在しない偽遺伝子。偽遺伝子が親遺伝子であり、一部の生物種では機能するが、別の生物種または系統ではなんらかの方法（インデル、変異、組み換え）で破壊されている元の遺伝子。  対立遺伝子(allelic) – ある個体群に常に認められる遺伝子であるが、その個体群は機能的に異なる対立遺伝子も有している。すなわち、ある系統では機能が備わっている遺伝子を有しているが、別の系統は偽遺伝子を有しているという場合もある。主要組織適合抗原複合体(MHC)ハプロタイプには、対立遺伝子の偽遺伝子が存在する。  不明(unknown) - 出願人は、偽遺伝子化の方法を知らない。
6.55.	qualifier 定義 qualifier 値のフォーマット コメント	rearranged エントリで記載された配列が、適応免疫反応の一部として体細胞で再配列されたものであり、親の生殖細胞から受け継いだ再配列されていない配列ではないこと。 なし qualifier 「rearranged」は、適応免疫反応に関与しない染色体再配列の注釈に使用してはならない；qualifier 「germline」および qualifier 「rearranged」は、同じソース(source) feature で使用してはならない。qualifier 「germline」および qualifier 「rearranged」は、適応免疫反応の一部として体細胞再配列を起こす可能性のある分子にのみ使用しなければならない。これ

らは、頸のある脊椎動物の t-細胞受容体 (TCR) および免疫グロブリン遺伝子座、および頸のない魚 (ヤツメウナギとメクラウナギ) の無関係な可変リンパ球受容体 (VLR) 遺伝子座である；qualifier 「germline」 および qualifier 「rearranged」 は有頭動物(系統分類 ID = 89593)以外では使用してはならない。

---

6.56.	qualifier	recombination_class
	定義	配列内の組換えホットスポット領域の分類を構造化して記述したもの。
	必須 qualifier 値の フォーマット	TYPE 上記の TYPE は以下の統制語彙または統制句の 1 つである。 meiotic mitotic non_allelic_homologous chromosome_breakpoint other
	例	<INSDQualifier_value>meiotic</INSDQualifier_value> <INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value>
	コメント	統制語彙 recombination_class に用いられていない特定の組換えクラスは、qualifier 「recombination_class」 の qualifier 値として「other」を入力し、新規の組換えクラスに関する簡単な説明を qualifier 「note」 に記述することにより、注釈を付さなければならない。
6.57.	qualifier	regulatory_class
	定義	配列内の転写、翻訳、複製、染色体構造に関連する調節エレメントの分類を構造的に記述したもの。
	必須 qualifier 値の フォーマット	TYPE 上記の TYPE は以下の統制語彙または統制句の 1 つである。 attenuator CAAT_signal DNase_I_hypersensitive_site enhancer enhancer_blocking_element GC_signal imprinting_control_region insulator locus_control_region matrix_attachment_region minus_35_signal minus_10_signal polyA_signal_sequence promoter recoding_stimulatory_region recombination_enhancer replication_regulatory_region response_element ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region uORF other
	例	<INSDQualifier_value>promoter</INSDQualifier_value> <INSDQualifier_value>enhancer</INSDQualifier_value> <INSDQualifier_value>ribosome_binding_site</INSDQualifier_value>
	コメント	統制語彙 regulatory_class にまだ記載されていない特定の規制クラスは、qualifier 「regulatory_class」 の値として「other」を入力し、新規の規制クラスについての簡単な説明を qualifier 「note」 に記述することで注釈を付さなければならない。
6.58.	qualifier	replace

---

	<b>定義</b>	feature の location で特定された配列が、qualifier 「replace」の qualifier 値で示される配列に置き換えられることを示す；Qualifier に配列が記載されていない（すなわち、qualifier 値が記載されていない）場合、欠失を意味する。
	<b>必須 qualifier 値のフォーマット</b>	フリーテキスト
	<b>例</b>	<INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value></INSDQualifier_value> - 欠失の場合
6.59.	<b>qualifier</b>	<b>ribosomal_slippage</b>
	<b>定義</b>	蛋白質の翻訳中に、特定の配列がリボソームにプログラムされ、リボソームスリップと呼ばれるメカニズムにより、代替のリーディングフレームに変更されることがある。
	<b>qualifier 値のフォーマット</b>	なし
	<b>コメント</b>	feature 「CDS」の位置には、例えば[join(486..1784,1787..4810)]のような存在位置演算子を使用して、ribosomal_slippage の location を示す必要がある。
6.60.	<b>qualifier</b>	<b>rpt_family</b>
	<b>定義</b>	反復配列の種類、例えば制限酵素アルスロバクター・ルテウス(Alu)またはクレブシエラ・ニューモニエ(Kpn)。
	<b>必須 qualifier 値のフォーマット</b>	フリーテキスト
	<b>例</b>	<INSDQualifier_value>Alu</INSDQualifier_value>
6.61.	<b>qualifier</b>	<b>rpt_type</b>
	<b>定義</b>	反復配列の構造と分布
	<b>必須 qualifier 値のフォーマット</b>	以下の統制語彙または統制句の 1 つを qualifier 値に用いる。
		tandem
		direct
		inverted
		flanking
		nested
		terminal
		dispersed
		long_terminal_repeat
		non_ltr_retrotransposon_polymeric_tract
		centromeric_repeat
		telomeric_repeat
		x_element_combinatorial_repeat
		y_prime_element
		other
	<b>例</b>	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
	<b>コメント</b>	上述の qualifier 値の定義は以下の通りである。
		直列(tandem) - 同一の配列が隣り合って同じ方向で繰り返し配置された配列。
		直接(direct) - 同一の配列が必ずしも隣り合って存在しないが、同じ方向で繰り返し配置された配列。
		反転(inverted) - 1 つの分子上で同一の配列が逆方向に繰り返し配置された配列。
		隣接(flanking) - 機能上の重要な影響を及ぼす配列の外に存在する反復配列(例 トランスポゾンの挿入部位)
		入れ子(nested) - 他の要素の挿入により中断されているが、繰り返し配置された配列。
		分散(dispersed) - 遺伝子全体に散在している反復配列。
		末端(terminal) - 機能上の重要な影響を及ぼす配列上の末端で、繰り返し配置された配列(例 末端に長い反復配列を有する(LTR型)トランスポゾン)

**long\_terminal\_repeat** - 定義された配列の両端で、同じ方向に繰り返し配置された配列。一般的にレトロウイルスの配列上で認められる種類の反復配列。

**non\_ltr\_retrotransposon\_polymeric tract** - 非 LTR 型レトロトランスポゾンの配列上のポリ(dA)等の、ポリトラクト。

**centromeric\_repeat** - セントロメアの反復配列の中で認められる、反復配列。

**telomeric\_repeat** - テロメアの配列の中で認められる、反復配列。

**x\_element\_combinatorial\_repeat** - X エレメントとテロメア、またはその隣接する Y' エлементの間に存在する反復配列。

**y\_prime\_element** - 1 コピーまたは 2 から 4 コピーの縦列反復であるかいずれか一方の、テロメア反復配列または、X エレメントのコンビナトリアル反復配列に隣接して配置された反復配列の領域。

**その他(other)** – 他の値では表現できない重要な属性を示す反復配列。

6.62.	qualifier	rpt_unit_range
	定義	範囲で表現された反復単位の location。
	必須 qualifier 値の フォーマット	<base_range> 上述の<base_range>には、反復ユニットの先頭の塩基の位置と末端の塩基の位置が 2 つのドットで区切られて示されている。
	例	<INSDQualifier_value>202..245</INSDQualifier_value>
	コメント	feature key 「oriT」 および 「repeat_region」 で指定された領域内で、反復単位を構成する配列の塩基の範囲を示すために使用する。
6.63.	qualifier	rpt_unit_seq
	定義	反復配列の同一性。
	必須 qualifier 値の フォーマット	フリーテキスト
	例	<INSDQualifier_value>aagggc</INSDQualifier_value> <INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value> <INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value>
	コメント	feature key 「oriT」 および 「repeat_region」 で指定された領域内の反復単位を構成する文字通りの配列を示すために使用する。
6.64	qualifier	satellite
	定義	サテライト DNA マーカーの識別子であり、短い基本反復単位の多数のタンデム反復（同一または関連）で構成される。
	必須 qualifier 値の フォーマット	<satellite_type>[<class>][<identifier>] 上記の<satellite_type>は以下のいずれかの 1 つである。 satellite; microsatellite; minisatellite
	例	<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>
	コメント	多くのサテライトは、塩基組成やその他の特性が他のゲノムと異なるため、それらを識別することができる。
6.65.	qualifier	segment
	定義	配列されたウイルスまたはファージのセグメントの名称。
	必須 qualifier 値の フォーマット	フリーテキスト
	例	<INSDQualifier_value>6</INSDQualifier_value>

---

6.66.	qualifier	serotype 抗原性を特徴とする種の血清学的多様性。
	定義	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	必須 qualifier 値の フォーマット	<INSDQualifier_value>B1</INSDQualifier_value>
	例	feature key 「source」と共にのみ用いられる；国際原核生物命名規約(bacteriological Code)では、原核生物には「血清型」ではなく「血液型亜型」という用語を使用することを推奨している；国際原核生物命名規約(1990 年版)、附属書 10.8 、「種内用語」を参照のこと。
	コメント	
6.67.	qualifier	serovar 抗原性を特徴とする種（通常は原核生物）の血清学的変異体
	定義	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	必須 qualifier 値の フォーマット	<INSDQualifier_value>O157:H7</INSDQualifier_value>
	例	feature key 「source」と共にのみ用いられる；国際原核生物命名規約では、原核生物に対して「serotype」の代わりに「serovar」という用語を使用することを推奨している。
	コメント	
6.68.	qualifier	sex 配列が得られた生物の性別； 真核生物で減数分裂を行い、性的に異なる配偶子を持つ場合は性別が用いられる。
	定義	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	必須 qualifier 値の フォーマット	<INSDQualifier_value>female</INSDQualifier_value>
	例	<INSDQualifier_value>male</INSDQualifier_value>
	コメント	<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] qualifier 「sex」 (qualifier 「mating_type」 ではなく)は、後生動物、有胚植物類、紅藻類＆褐藻類に関して用いるべきである；細菌、古細菌&菌類界に関しては、qualifier 「mating_type」 (qualifier 「sex」 ではなく)を用いるべきである；ウイルスに関しては、qualifier 「sex」 および qualifier 「mating_type」 は用いるべきではない；その他(上述の分類群以外)の詳細情報を記載する場合で、上の例で用いられている用語を qualifier 「sex」 の qualifier 値として用いない場合は、qualifier 「mating_type」 を用いるべきである。
6.69.	qualifier	standard_name この feature の標準的な名称。
	定義	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	必須 qualifier 値の フォーマット	<INSDQualifier_value>dotted</INSDQualifier_value>
	例	
	コメント	遺伝子名を省略せずに記載する場合は qualifier 「standard_name」 を使用し、qualifier 「gene」 を用いて遺伝子コードを与える (例えば、上の例では、qualifier 「gene」 の値は 「Dt」 )。
6.70.	qualifier	strain 配列が得られた strain
	定義	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	必須 qualifier 値の フォーマット	<INSDQualifier_value>BALB/c</INSDQualifier_value>
	例	
	コメント	qualifier 「strain」 を記載する feature エントリには、qualifier 「environmental_sample」 を用いてはならない。
6.71.	qualifier	sub_clone 配列が得られたサブクローン
	定義	フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。
	必須 qualifier 値の フォーマット	<INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value>
	例	

---

コメント		
ソース(source) feature は、2つ以上の qualifier 「sub_clone」 を用いてはならない；複数のサブクローニングから得られた配列であることを示すために、複数のソース(source)を feature key 「misc_feature」 およびその qualifier 「note」 を用いて詳細情報を記載しても良い。		
6.72.	qualifier 定義 必須 qualifier 値の フォーマット 例	sub_species 配列が得られた生物の亜種の名称。 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>lactis</INSDQualifier_value>
6.73.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	sub_strain 親 strain (qualifier 「strain」 に注釈を付すべき) から派生した、配列が得られた遺伝子的またはその他の方法で改変された sub_strain の名称または識別子。 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>abis</INSDQualifier_value> ソース(source) feature の qualifier 「strain」 と共に用いられなければならない；親の系統が与えられていない場合は、qualifier 「sub_strain」 ではなく qualifier 「strain」 を用いて亜系統に注釈を付すべきである。例えば、qualifier 「strain」 (qualifier 値は K-12)、qualifier 「sub_strain」 (qualifier 値は MG1655) または、qualifier 「strain」 (qualifier 値は MG1655) のいずれかである。
6.74.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	tag_peptide tmRNA のタンパク質分解タグ用ポリペプチドをコードする塩基位置とその終止コドン。 <base_range> 上述の<base_range>には、蛋白質分解のタグが存在する配列上の先頭の塩基の位置と末端の塩基の位置が 2 つのドットで区切られて示されている。 <INSDQualifier_value>90..122</INSDQualifier_value> タグ_ペプチドに対応するアミノ酸配列は、5' 末端の部分的な feature 「CDS」 を示すものとして注釈を付すべきことを推奨する；例：<90...122> の location にある CDS。
6.75.	qualifier 定義 必須 qualifier 値の フォーマット 例	tissue_lib 配列が得られた組織ライブラリ。 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>tissue library 772</INSDQualifier_value>
6.76.	qualifier 定義 必須 qualifier 値の フォーマット 例	tissue_type 配列が得られた組織のタイプ フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>liver</INSDQualifier_value>
6.77.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	transl_except 翻訳上の例外；ある 1 つのコドンの翻訳が、qualifier 「organism」 または qualifier 「transl_table」 に記載されている遺伝コードと一致しない状態。 (pos: <location,aa>:<amino_acid>) ここで、<amino_acid>は、base_range の位置にあるコドンによってコードされるアミノ酸の 3 文字の略語である。 <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> アミノ酸が本附属書の第 3 節に記載されている特定のアミノ酸でない場合、「OTHER」を<amino_acid>として使用し、例外的なアミノ酸の名称を qualifier 「note」 で記載する；修飾アミノ酸であるセレノシスティンについては、<amino_acid>に 3 文字の略語「Sec」(アミノ酸配列内の 1 文字の記号「U」)を使用し、修飾アミノ酸であるピロリジンについては、<amino_acid>に 3 文字の略語「Pyl」(アミノ酸配列内の 1 文字の記号「O」)を使用する； TAA 終止コドンが mRNA に 3'A 残基を付加することで完成する部分終止コドンの場合、その location には単一の base_position または base_range のいずれかを使用し、「mRNA に 3'A 残

基を付加することで完成する終止コドン」を示す qualifier 「note」と併せて用いる（上記の 3 番目と 4 番目の例を参照されたい）。

---

6.78.	qualifier 定義 必須 qualifier 値のフォーマット 例 コメント	transl_table 普遍的または標準的な遺伝コード表以外の場合に使用される遺伝コード表の定義。使用されるテーブルは本附属書に記載がある。 <integer> 上記の<integer>とは、遺伝コード表に割り振られた番号である。 <INSDQualifier_value>3</INSDQualifier_value> 上記の例では、酵母のミトコンドリアの遺伝コード表を用いて配列表を作成している。 feature key 「CDS」の詳細情報が qualifier 「transl_table」に記載されていない場合は、CDS は標準コード（すなわち Universal Genetic Code）を使用して翻訳されます。指定されたテーブルの範囲外の遺伝子コードの例外は qualifier 「transl_except」で報告される。
6.79.	qualifier 定義 qualifier 値のフォーマット 例 コメント	trans_splicing 2 つの RNA 分子のエクソンが連結し、成熟 RNA を形成することを示している。 なし トランススプライシングの結果產生された feature（例えば CDS、mRNA 等）の詳細情報を記載するために本 qualifier を用いるべきである。またこの qualifier は、スプライシングの発生が feature location の演算子「joint」、例 joint (complement (69611..69724),139856..140087) を用いて示されている場合にのみ使用されなければならない。
6.80.	qualifier 定義 必須 qualifier 値のフォーマット 例 コメント	translation 標準（またはユニバーサル）遺伝コード、または qualifier 「transl_table」で指定された遺伝コード表や qualifier 「transl_except」の例外によって決められた遺伝コード表のいずれかに準じた 1 文字の略号によるアミノ酸配列。 本附属書の第 3 節に記載されている 1 文字のアミノ酸の記号を用いて記載した、連続した 1 本のアミノ酸配列。記号「X」は AA 等の例外的なアミノ酸に用いる記号である。 <INSDQualifier_value>MASTFPPWYRGCASTPSLKGLIMCTW</INSDQualifier_value> CDS フィーチャーと共にのみ使用される；翻訳産物に 4 つ以上の特定のアミノ酸が含まれる場合、qualifier 「protein_id」を付さなければならぬ；遺伝コード表の定義および location について transl_table を参照されたい；feature 「CDS」を追加で注釈を付すには、translation qualifier、「pseudo」、「pseudogene」のいずれかのみが許容されている。
6.81.	qualifier 定義 必須 qualifier 値のフォーマット 例 コメント	variety 配列の元となった生物の種類（＝変種、正式なリンネ式分類法の階級）  言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>insularis</INSDQualifier_value> 植物の栽培品種（すなわち人為的な選択により栽培された植物）を記載するには qualifier 「cultivar」を用いる；植物および菌類の変種以外の変種は、qualifier 「note」を用いて注釈を付すべきであり、その際、qualifier 値として例えば次を用いる： <INSDQualifier_value>breed:Cukorova</INSDQualifier_value>。

---

## 第7節：アミノ酸配列のfeature keys

ここでは、アミノ酸配列に使用できる feature key の一覧を示している。 feature key はアルファベット順に並んでいる。

7.1.	feature key 定義 任意の qualifier コメント	ACT_SITE 酵素の活性に関与するアミノ酸 qualifier「note」 活性部位に存在するアミノ酸残基は、feature key 「ACT_SITE」を用いて個別に注釈を付さなければならない。また対応するアミノ酸残基の番号を、feature location の要素に存在位置記述子として記載しなければならない。
7.2.	feature key 定義 必須 qualifier コメント	BINDING 任意の化学基(補酵素、補欠分子族等)の結合部位。基の化学的な性質は qualifier 「note」を用いて記載する。 qualifier 「note」 qualifier 「note」の値の例としては、「Heme(covalent)」および「Chloride」がある。必要に応じて feature key 「BINDING」ではなく、feature key 「CA_BIND」、feature key 「DNA_BIND」、feature key 「METAL」および feature key 「NP_BIND」を用いるべきである。
7.3.	feature key 定義 任意の qualifier	CA_BIND カルシウム結合の領域の範囲 qualifier 「note」
7.4.	feature key 定義 必須 qualifier コメント	CARBOHYD グリコシリ化した部位 qualifier 「note」 本 key は、糖類（単糖または多糖）が蛋白質の残基に付加していることを示している。この様な種類の蛋白質への付加(C-, N- または O-結合グルコシリ化)は、qualifier 「note」に記載する。還元末端糖の性質が分かっている場合は、その略語を括弧の間に記載する。還元末端糖の略語の後に 3 つのドット「...」が続く場合は、糖鎖がさらに続いていることを示している。それとは逆に、ドットが 1 つも付いていない場合は单糖類が結合していることを意味している。qualifier 「note」の qualifier 値の例には、次がある；N-結合(GlcNAc...)；O-結合(GlcNAc)；O-結合(Glc...)；C-結合(Man)部分；O-結合(Ara...)。
7.5.	feature key 定義 任意の qualifier	CHAIN 成熟蛋白質のポリペプチド鎖の長さ qualifier 「note」
7.6.	feature key 定義 任意の qualifier	COILED 二重コイル構造を有する領域の範囲 qualifier 「note」
7.7.	feature key 定義 任意の qualifier	COMPIAS compositionally biased region(CBR。組成の偏った領域) qualifier 「note」
7.8.	feature key 定義 任意の qualifier コメント	CONFLICT 異なるソースが異なる配列を報告すること qualifier 「note」 qualifier 「note」の qualifier 値の例；Missing; K -> Q; GSDSE -> RIRLR; V -> A.
7.9.	feature key 定義 必須 qualifier コメント	CROSSLNK 翻訳語に形成されたアミノ酸同士の結合 qualifier 「note」 2 つの蛋白質の間（鎖間架橋）、または同じ蛋白質の 2 つの部分の間（鎖内架橋）に形成される様々なタイプの共有結合。但しジスフィルド結合により形成される架橋には、feature key 「DISULFID」が使用される。鎖間架橋の場合、feature location 要素の存在位置記述子は、他の蛋白質に架橋されたアミノ酸の残基番号となる。鎖内架橋の場合、feature location

要素の存在位置記述子は、架橋されたアミノ酸の残基番号を「x..y」形式で表したもので、例えば、「42...50」のようになる。qualifier「note」は、架橋の性質を示すもので、少なくとも結合体の名前と関係する2つのアミノ酸の同一性を明示する。qualifier「note」の値の例。「イソグルタミルシステインチオエステル (Cys-Gln)」、「 $\beta$ -メチルランチオニン (Cys-Thr)」、「グリシルリジンイソペプチド (Lys-Gly)」(ユビキチンの G-Cter との鎖間)」。

---

7.10.	feature key 定義 必須 qualifier コメント	DISULFID ジスフィルド結合 qualifier「note」 鎖間ジスルフィド結合の場合、feature location の要素の存在位置記述子は、他の蛋白質に結合しているシステインの残基番号となる。鎖内架橋の場合、feature location の要素の存在位置記述子は、リンクされたシステインの残基番号を「x..y」形式で表したもので、例えば「42...50」のようになる。鎖間ジスルフィド結合の場合、note qualifier は、他の蛋白質を特定することで、架橋の性質を示す。
7.11.	feature key 定義 必須 qualifier コメント	DNA_BIND DNA 結合領域の範囲 qualifier「note」 DNA 結合領域の性質は、qualifier「note」を用いて記載する。その際の qualifier 値の例としては、qualifier「note」の例として、「Homeobox」および「Myb 2」がある。
7.12.	feature key 定義 必須 qualifier コメント	DOMAIN 特徴的な三次元構造または折り畳まれた構造を有する、二次構造の特殊な組み合わせとして定義されるドメインの範囲。 qualifier「note」 ドメインのタイプは qualifier「note」を用いて記載する。ドメインのコピーが複数存在する場合は番号を付与する。この qualifier「note」の qualifier 値の例としては、「Ras-GAP」および「Cadherin1」がある。
7.13.	feature key 定義 任意の qualifier コメント	HELIX 二次構造、すなわちヘリックス（例 $\alpha$ ヘリックス、3(10)ヘリックスまたは Pi ヘリックス（TT ヘリックス） qualifier「note」 feature key「HELIX」は三次構造が判明している蛋白質にのみ用いる。二次構造は 3 種類のみ指定されている（ヘリックス [key「HELIX」]、 $\beta$ ストランド [key「STRAND」] および ターン [key「TURN」]）。これらの種類の 1 つに該当しない残基は、「look」または「random-coil」構造である。
7.14.	feature key 定義 任意の qualifier コメント	INIT_MET イニシエーターメチオニン qualifier「note」 feature location の要素の存在位置記述子は「1」である。本 feature key は、N 末端のメチオニンが切断されていることを示している。イニシエーターメチオニンが切断されていない場合は使用されない。
7.15.	feature key 定義 任意の qualifier	INTRAMEM 細胞膜に位置する領域の範囲で、細胞膜を横切らないもの。 qualifier「note」
7.16.	feature key 定義 必須 qualifier コメント	LIPID 脂質の部分との共有結合 qualifier「note」 共有結合している脂質の部分の化学的性質は、qualifier「note」を用いて記載するが、少なくとも脂質修飾を受けたアミノ酸の名称を記載する。qualifier「note」の値の例には、「N-myristoyl glycine」、「GPI-anchor amidated serine」、「S-diacylglycerol cysteine」など。
7.17.	feature key 定義 必須 qualifier	METAL 金属イオンの結合部位 qualifier「note」

---

	<b>コメント</b>	qualifier 「note」 は金属の性質を示す。qualifier 「note」 の qualifier 値の例には、「Iron (heme axial ligand)」 および 「Copper」 がある。
7.18.	feature key 定義 必須 qualifier コメント	MOD_RES 残基の翻訳後修飾 qualifier 「note」 修飾を受けた残基の化学的性質は qualifier 「note」 を用いて記載するが、少なくとも翻訳後修飾を受けたアミノ酸の名称を示す。修飾アミノ酸が本附属書の第 4 節に記載されている場合は、完全な名称の代わりに略語を使用しても良い。qualifier 「note」 の qualifier 値の例には、「3-Hyp」 および 「MeLys」 および 「N-6-methyllysine」 がある。
7.19.	feature key 定義 任意の qualifier	MOTIF 生物学的に関心のある、短い配列モチーフ(最大で 20 個のアミノ酸が連続した配列) qualifier 「note」
7.20.	feature key 定義 任意の qualifier	MUTAGEN 変異により実験的に変化した部位 qualifier 「note」
7.21.	feature key 定義 任意の qualifier コメント	NON_STD 非標準アミノ酸 qualifier 「note」 この key は、アミノ酸配列内の非標準アミノ酸であるセレノシステイン(U)およびピロリシン(O)の出現率のみを表している。
7.22.	feature key 定義 任意の qualifier コメント	NON_TER 配列の端にある残基は末端の残基ではないこと。 qualifier 「note」 配列の先頭に適用する場合、これは最初の位置が完全な分子の N-末端ではないことを意味している。最後の位置に適用する場合、この位置が完全な分子の C-末端ではないことを意味している。
7.23.	feature key 定義 必須 qualifier コメント	NP_BIND ヌクレオチドのリン酸結合領域の範囲 qualifier 「note」 ヌクレオチドのリン酸の性質は、qualifier 「note」 を用いて記載する。その際の qualifier 値の例には、「ATP」 および 「FAD」 がある。
7.24.	feature key 定義 任意の qualifier	PEPTIDE 放出された活性ポリペプチドの範囲 qualifier 「note」
7.25.	feature key 定義 任意の qualifier	PROPEP プロペプチドの範囲 qualifier 「note」
7.26.	feature key 定義 任意の qualifier	REGION 配列内の関心領域の範囲 qualifier 「note」
7.27.	feature key 定義 任意の qualifier	REPEAT 内部配列の反復の範囲 qualifier 「note」
7.28.	feature key 定義 任意の qualifier	SIGNAL シグナル配列の長さ (プレペプチド) qualifier 「note」
7.29.	feature key	SITE

	<b>定義</b>	他の feature key で定義されていない、配列上の関心のある単一のアミノ酸部位。2つの隣接するアミノ酸の位置で示される、アミノ酸の結合にも適用できる。
	<b>必須 qualifier</b>	qualifier 「note」
	<b>コメント</b>	修飾アミノ酸の注釈に使用される場合、qualifier 「note」 の値は、本附属書の第 4 節で示した略語か、もしくは完全で、省略されていない修飾アミノ酸の名称のいずれかでなければならない。
7.30.	<b>feature key</b>	source
	<b>定義</b>	その配列の起源(source)を示す。この key は必須である。全ての配列には、配列全体にまたがる単一の feature 「source」 がある。
	<b>必須 qualifier</b>	mol_type
	<b>任意の qualifier</b>	organism qualifier 「note」
7.31.	<b>featurekey</b>	STRAND
	<b>定義</b>	2 次構造；βストランド；例えば水素結合した βストランドまたは孤立した βブリッジの残基
	<b>任意の qualifier</b>	qualifier 「note」
	<b>コメント</b>	この key は、三次構造が判明している蛋白質にのみ使用する。二次構造は 3 種類のみに指定されている（ヘリックス構造 [key HELIX]、βストランド構造 [key STRAND] およびターン構造 [key TURN]）。これらのクラスのいずれかに該当しない残基は、「ループ」または「ランダムコイル」構造である。
7.32.	<b>feature key</b>	TOPO_DOM
	<b>定義</b>	トポジカルドメイン
	<b>任意の qualifier</b>	qualifier 「note」
7.33.	<b>feature key</b>	TRANSMEM
	<b>定義</b>	膜貫通領域の範囲
	<b>任意の qualifier</b>	qualifier 「note」
7.34.	<b>feature key</b>	TRANSIT
	<b>定義</b>	トランジットペプチドの範囲（ミトコンドリア、葉緑体、チラコイド、シアネル、ペルオキシソーム等）
	<b>任意の qualifier</b>	qualifier 「note」
7.35.	<b>feature key</b>	TURN
	<b>定義</b>	二次構造のターン。例えば、水素結合したターン（3-ターン、4-ターンまたは 5-ターン）
	<b>任意の qualifier</b>	qualifier 「note」
	<b>コメント</b>	この feature は、三次構造が判明している蛋白質にのみ使用する。二次構造は 3 種類のみに指定されている（ヘリックス構造 [key HELIX]、βストランド構造 [key STRAND] およびターン構造 [key TURN]）。これらのクラスのいずれかに該当しない残基は、「ループ」または「ランダムコイル」構造である。
7.36.	<b>feature key</b>	UNSURE
	<b>定義</b>	配列内の不確実な点
	<b>任意の qualifier</b>	qualifier 「note」
	<b>コメント</b>	アミノ酸配列内で、著者が配列表示について確信が持てない領域を記述するために仕様する。
7.37.	<b>feature key</b>	VARIANT
	<b>定義</b>	配列内の変異体の存在を著者が報告する
	<b>任意の qualifier</b>	qualifier 「note」
7.38.	<b>feature key</b>	VAR_SEQ
	<b>定義</b>	選択的スプライシング、代替プロモーターの使用、代替開始およびリボソームフレームシフトによって生じた配列変異体の説明
	<b>任意の qualifier</b>	qualifier 「note」
7.39.	<b>feature key</b>	ZN_FING

---

定義	ジンクフィンガー領域の範囲
任意の qualifier	qualifier 「note」
コメント	ジンクフィンガーのタイプは、qualifier 「note」 に記載する。例えば、「GATA-type」、「NR C4-type」がある。

---

## 第 8 節：アミノ酸配列のqualifiers

本節では、アミノ酸配列に使用できる qualifier の一覧を示している。

qualifier 値のフォーマットが言語に依存するフリーテキストの場合、以下のいずれかを使用しなければならない。

- 1) 要素 INSDQualifier\_value、または
- 2) 要素 NonEnglishQualifier\_value、または
- 3) 要素 INSDQualifier\_value と要素NonEnglishQualifier\_value の両方。

値のフォーマットが言語依存のフリーテキストとして識別されない場合、要素 INSDQualifier\_value を使用しなければならず、要素 NonEnglishQualifier\_value は使用してはならない。

注意：言語に依存するフリーテキスト値形式を有する qualifier に対して提供される qualifier 値は、国際的、国内的、または広域的な手続きのために翻訳が必要な場合がある。次の表に示す qualifier は、言語に依存したフリーテキストの値を持つとみなされる。

表 6: 言語に依存したフリーテキスト値を持つアミノ酸配列の qualifier の一覧

節	言語に依存したフリーテキスト qualifier
8.2	note
8.3	organism

8.1.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	mol_type 配列の生体内における分子型 protein <INSDQualifier_value>protein</INSDQualifier_value> feature key source の必須 qualifier は 「mol_type」。
8.2.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	note コメントおよび追記のすべて フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Heme (covalent)</INSDQualifier_value> 以下の feature key の必須 qualifier は qualifier 「note」である。 feature key 「BINDING」, feature key 「CARBOHYD」, featurekey 「CROSSLNK」, feature key 「DISULFID」, feature key 「DNA_BIND」, feature key 「DOMAIN」, feature key 「LIPID」, feature key 「METAL」, feature key 「MOD_RES」, feature key 「NP_BIND」, feature key 「SITE」および、feature key 「ZN_FING」。
8.3.	qualifier 定義 必須 qualifier 値の フォーマット 例 コメント	organism ペプチドを有する生物の学名 フリーテキスト 言語依存：この値は、国際的／国内的／広域的な手続きのために翻訳が必要な場合がある。 <INSDQualifier_value>Homo sapiens</INSDQualifier_value> qualifier 「organism」は、feature key 「source」の必須 qualifier である。

## 第9節：遺伝コード表

表7は、コードされた配列を翻訳するために使用される遺伝コード表を再現している。qualifier「trans\_table」の値は、対応する遺伝コード表に割り当てられた番号である。feature「CDS」にtranslation qualifierが記載されていてもqualifier「transl\_table」が記載されていない場合は、デフォルトで1-Standard Codeが翻訳に使用される(注：遺伝コード表7、8、15、17から20は存在しないため、これらの番号は表7には記載されていない)。

表7：遺伝コード表

1 - Standard Code	
AAs =	FFLLSSSSYY**CC*WLLLLPPPPHQRRRIIIMTTTNNKSSRRVVVVAAAADDEEGGGG..
Starts =	--M-----M-----M-----..
Base1 =	tttttttttttttccccccccccccccaaaaaaaaggggggggggggggggg..
Base2 =	tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.	.
2 - Vertebrate Mitochondrial Code	
AAs =	FFLLSSSSYY**CCWWLLLLPPPPHQRRRIIIMTTTNNKSSRRVVVVAAAADDEEGGGG..
Starts =	-----MMMM-----M-----..
Base1 =	ttttttttttttttccccccccccccccaaaaaaaaggggggggggggggggg..
Base2 =	tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.	.
3 - Yeast Mitochondrial Code	
AAs =	FFLLSSSSYY**CCWWTTTTPPPPHQRRRIIIMTTTNNKSSRRVVVVAAAADDEEGGGG..
Starts =	-----MM-----M-----..
Base1 =	ttttttttttttttccccccccccccccaaaaaaaaggggggggggggggggg..
Base2 =	tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.	.
4 - Mold, Protozoan, Coelenterate Mitochondrial Code & Mycoplasma/Spiroplasma Code	
AAs =	FFLLSSSSYY**CCWWLLLLPPPPHQRRRIIIMTTTNNKSSRRVVVVAAAADDEEGGGG..
Starts =	--MM-----M-----MMMM-----M-----..
Base1 =	ttttttttttttttccccccccccccccaaaaaaaaggggggggggggggggg..
Base2 =	tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.	.
5 - Invertebrate Mitochondrial Code	
AAs =	FFLLSSSSYY**CCWWLLLLPPPPHQRRRIIIMTTTNNKSSSSVVVVAAAADDEEGGGG..
Starts =	--M-----MMMM-----M-----..
Base1 =	ttttttttttttttccccccccccccccaaaaaaaaggggggggggggggggg..
Base2 =	tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.	.
6 - Ciliate, Dasycladacean and Hexamita Nuclear Code	
AAs =	FFLLSSSSYYQQCC*WLLLLPPPPHQRRRIIIMTTTNNKSSRRVVVVAAAADDEEGGGG..
Starts =	-----M-----..
Base1 =	ttttttttttttttccccccccccccccaaaaaaaaggggggggggggggg..
Base2 =	tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.	.

**9 - Echinoderm and Flatworm Mitochondrial Code**

```

AAS = FFLLSSSSYY**CCWWLLLLPPPPHQRRRIIIMTTTNNNKSSSSVVVAAAADDEEGGGG..
Starts = -----M-----M-----
Base1 = tttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttcccaaaagggtttccccaaaagggtttccccaaaagggtttccccaaaagg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc..
.
```

**10 - Euplotid Nuclear Code**

```

AAS = FFLLSSSSYY**CCCWLLLLPPPPHQRRRIIIMTTTNNKSSRRVVVAAAADDEEGGGG..
Starts = -----M-----M-----
Base1 = tttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttcccaaaagggtttccccaaaagggtttccccaaaagggtttccccaaaagg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc..
.
```

**11 - Bacterial, Archaeal, and Plant Plastid Code**

```

AAS = FFLLSSSSYY**CC=WLLLLPPPPHQRRRIIIMTTTNNKSSRRVVVAAAADDEEGGGG..
Starts = --M-----M-----MMMM-----M-----
Base1 = tttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttcccaaaagggtttccccaaaagggtttccccaaaagggtttccccaaaagg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc..
.
```

**12 - Alternative Yeast Nuclear Code**

```

AAS = FFLLSSSSYY**CC=WLLLSPPPPHQRRRIIIMTTTNNKSSRRVVVAAAADDEEGGGG..
Starts = -----M-----M-----
Base1 = tttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttcccaaaagggtttccccaaaagggtttccccaaaagggtttccccaaaagg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc..
.
```

**13 - Ascidian Mitochondrial Code**

```

AAS = FFLLSSSSYY**CCWWLLLLPPPPHQRRRIIIMMTTNNKSSGGVVVAAAADDEEGGGG..
Starts = --M-----MM-----M-----
Base1 = tttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttcccaaaagggtttccccaaaagggtttccccaaaagggtttccccaaaagg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc..
.
```

**14 - Alternative Flatworm Mitochondrial Code**

```

AAS = FFLLSSSSYY*CCWWLLLLPPPPHQRRRIIIMTTTNNKSSSSVVVAAAADDEEGGGG..
Starts = -----M-----
Base1 = tttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttcccaaaagggtttccccaaaagggtttccccaaaagggtttccccaaaagg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc..
.
```

**16 - Chlorophycean Mitochondrial Code**

```

AAS = FFLLSSSSYY*LCC=WLLLSPPPPHQRRRIIIMTTTNNKSSRRVVVAAAADDEEGGGG..
Starts = -----M-----
Base1 = tttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttcccaaaagggtttccccaaaagggtttccccaaaagggtttccccaaaagg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc..
.
```

**21 - Trematode Mitochondrial Code<sup>+</sup>**

```

AAS = FFLLSSSSYY**CCWWLLLLPPPPHQQRRRIIIMTTTNNKSSSSVVVVAADDEEGGGG..
Starts = -----M-----M-----
Base1 = tttttttttttttccccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.
```

**22 - Scenedesmus obliquus Mitochondrial Code<sup>+</sup>**

```

AAS = FFLLSS*SY*LC*WLLLLPPPPHQQRRRIIIMTTTNNKSSRRVVVVAADDEEGGGG..
Starts = -----M-----
Base1 = tttttttttttttccccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.
```

**23 - Thraustochytrium Mitochondrial Code<sup>+</sup>**

```

AAS = FF*LSSSSY*CC=WLLLLPPPPHQQRRRIIIMTTTNNKSSRRVVVVAADDEEGGGG..
Starts = -----M--M-----M-----
Base1 = tttttttttttttccccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.
```

**24 - Pterobranchia Mitochondrial Code<sup>+</sup>**

```

AAS = FFLLSSSSYY**CCWLLLLPPPPHQQRRRIIIMTTTNNKSSSKVVVVAADDEEGGGG..
Starts = ---M-----M-----M-----M-----
Base1 = tttttttttttttccccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.
```

**25 - Candidate Division SR1 and Gracilibacteria Code<sup>+</sup>**

```

AAS = FFLLSSSSYY**CCGWLLLLPPPPHQQRRRIIIMTTTNNKSSRRVVVVAADDEEGGGG..
Starts = ---M-----M-----M-----M-----
Base1 = tttttttttttttccccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.
```

**26 - Pachysolen tannophilus Nuclear Code<sup>+</sup>**

```

AAS = FFLLSSSSYY**CC=WLLLAPPPPQQRRRIIIMTTTNNKSSRRVVVVAADDEEGGGG..
Starts = -----M-----M-----
Base1 = tttttttttttttccccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.
```

**27 - Karyorelict Nuclear Code<sup>+</sup>**

```

AAS = FFLLSSSSYYQQCCWWLLLLPPPPHQQRRRIIIMTTTNNKSSRRVVVVAADDEEGGGG..
Starts = -----*-----M-----
Base1 = tttttttttttttccccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag..
.
```

**28 - Condyllostoma Nuclear Code**

```

AAs = FFLLSSSSYYQQCCWWLLLLPPPHHQRRRIIIMTTTNNKKSSRRVVVVAADDEEGGGG..
Starts = -----**-----M-----
Base1 = tttttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcag..
.
```

**29 - Mesodinium Nuclear Code**

```

AAs = FFLLSSSSYYYYCC=WLLLLPPPHHQRRRIIIMTTTNNKKSSRRVVVVAADDEEGGGG..
Starts = -----M-----
Base1 = tttttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcag..
.
```

**30 - Peritrich Nuclear Code**

```

AAs = FFLLSSSSYYEECC=WLLLLPPPHHQRRRIIIMTTTNNKKSSRRVVVVAADDEEGGGG..
Starts = -----M-----
Base1 = tttttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcag..
.
```

**31 - Blastocritidida Nuclear Code**

```

AAs = FFLLSSSSYYEECCWWLLLLPPPHHQRRRIIIMTTTNNKKSSRRVVVVAADDEEGGGG..
Starts = -----**-----M-----
Base1 = tttttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcag..
.
```

**33 - Cephalodiscidae Mitochondrial UAA-Tyr Code**

```

AAs = FFLLSSSSYYY=CCWWLLLLPPPHHQRRRIIIMTTTNNKKSSSKVVVVAADDEEGGGG..
Starts = ---M-----M-----M-----M-----
Base1 = tttttttttttttttccccccccccccaaaaaaaagggggggggggggggggg..
Base2 = tttccccaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg..
Base3 = tcagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcaagtcag..
.
```

[次頁から附属書Ⅱが続く]

## 附属書Ⅱ

## 配列表の文書型定義(DTD)

Version 1.3

2023年12月8日開催の第11回WIPO標準委員会(CWS)会合で採択

```

<?xml version="1.0" encoding="UTF-8"?>
<!--Annex II of WIPO Standard ST.26, Document Type Definition (DTD) for Sequence Listing

This entity may be identified by the PUBLIC identifier:
*****
PUBLIC "-//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
*****
* PUBLIC DTD URL

* https://www.wipo.int/standards/dtd/ST26SequenceListing_V1_3.dtd
*****

* Revision of Annex II to WIPO Standard ST.26 was approved by the Committee on WIPO
* Standards (CWS) at its tenth session.

*****
* CONTACTS
*****
xml.standards@wipo.int

*****
* NOTES
*****
The sequence data part is a subset of the complete INSDC DTD V.1.5 that only covers
the requirements of WIPO Standard ST.26.

*****
* REVISION HISTORY
*****
2022-11-25: Comment related to filename approved at CWS/10 (no update to version number)
2020-05-20: Version 1.3 approved at CWS/8.

Changes:
- Optional originalFreeTextLanguageCode attribute added to to allow applicants to indicate the
language of the free text in the original sequence listing.
- Optional nonEnglishFreeTextLanguageCode attribute added to to allow applicants to indicate the
language of the free text provided in the element .
- Optional id attribute added to INSDQualifier to facilitate comparison of language - dependent
qualifier values between sequence listings.
- Optional element added to element to allow applicants to type language-dependent qualifiers in a
non-English Language with the characters set forth in paragraph 40(a) of the ST.26 main body document.

2018-10-19: Version 1.2 approved at CWS/6.

Changes:
<INSDQualifier*> changed to <INSDQualifier+> for alignment with business needs and advice from NCBI
(an INSDFeature_quals element (if present) should have one or more INSDQualifier elements).

2017-06-02: Version 1.1 approved at the CWS/5

Changes:
Comments added to <INSDSeq_length>, <INSDSeq_division> and <INSDSeq_sequence> to clarify the reason of
the differences between the INSDC DTD v.1.5 and ST26 Sequence Listing DTD V1_1.

```

2016-03-24: Version 1.0 adopted at the CWS/4Bis

2014-03-11: Final draft for adoption.

\*\*\*\*\*

### ST26SequenceListing

\*\*\*\*\*

\* ROOT ELEMENT

\*\*\*\*\*

-->

```
<!ELEMENT ST26SequenceListing ((ApplicantFileReference | (ApplicationIdentification,
ApplicantFileReference?)), EarliestPriorityApplicationIdentification?, (ApplicantName,
ApplicantNameLatin?)?, (InventorName, InventorNameLatin?)?, InventionTitle+, SequenceTotalQuantity,
SequenceData+)>
```

<!--The elements ApplicantName and InventorName are optional in this DTD to facilitate  
the conversion between various encoding schemes-->

<!--originalFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single original language in  
which the language-dependent free text qualifiers (NonEnglishQualifier\_value) were prepared.

-->

<!--nonEnglishFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the language in which the  
language-dependent free text qualifiers (NonEnglishQualifier\_value) currently correspond.

-->

<!--fileName:

By default the file name will be set to the value provided for the project name in WIPO Sequence. If  
the value is identical to the actual ST.26 XML filename, it should be noted that Offices may enforce  
their requirements for the filename used which may restrict which characters are allowable for  
submitted electronic files. It is also acceptable for the value of the filename attribute and the  
actual file name to be different. Please refer to the WIPO Sequence and ST.26 Knowledge Base for  
further details on Offices' naming conventions for electronic files

-->

<!ATTLIST ST26SequenceListing

```
    dtdVersion CDATA #REQUIRED
    fileName CDATA #IMPLIED
    softwareName CDATA #IMPLIED
    softwareVersion CDATA #IMPLIED
    productionDate CDATA #IMPLIED
    originalFreeTextLanguageCode CDATA #IMPLIED
    nonEnglishFreeTextLanguageCode CDATA #IMPLIED
```

>

<!--ApplicantFileReference

Applicant's or agent's file reference, mandatory if application identification not provided.

-->

<!ELEMENT ApplicantFileReference (#PCDATA)>

<!--ApplicationIdentification

Application identification for which the sequence listing is submitted, when available.

-->

<!ELEMENT ApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--EarliestPriorityApplicationIdentification

Identification of the earliest priority application, which contains IPOFFICECODE,  
APPLICATIONNUMBERTEXT and FILINGDATE elements.

-->

<!ELEMENT EarliestPriorityApplicationIdentification (IPOfficeCode, ApplicationNumberText,
FilingDate?)>

<!--ApplicantName

The name of the first mentioned applicant in characters set forth in paragraph 40(a) of the ST.26 main  
body document.

-->

```

<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of names of
languages - Part 1: Alpha-2
-->
<!ELEMENT ApplicantName (#PCDATA)>
<!ATTLIST ApplicantName
    languageCode CDATA #REQUIRED

>
<!--ApplicantNameLatin
Where ApplicantName is typed in characters other than those as set forth in paragraph 40(b), a
translation or transliteration of the name of the first mentioned applicant must also be typed in
characters as set forth in paragraph 40(b) of the ST.26 main body document.
-->
<!ELEMENT ApplicantNameLatin (#PCDATA)>
<!--InventorName
Name of the first mentioned inventor typed in the characters as set forth in paragraph 40(a).-->
<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of names of
languages - Part 1: Alpha-2
-->
<!ELEMENT InventorName (#PCDATA)>
<!ATTLIST InventorName
    languageCode CDATA #REQUIRED

>
<!--InventorNameLatin
Where InventorName is typed in characters other than those as set forth in paragraph 40(b), a
translation or transliteration of the first mentioned inventor may also be typed in characters as set
forth in paragraph 40(b).
-->
<!ELEMENT InventorNameLatin (#PCDATA)>
<!--InventionTitle
Title of the invention typed in the characters as set forth in paragraph 40(a) in the language of
filing. A translation of the title of the invention into additional languages may be typed in the
characters as set forth in paragraph 40(a) using additional InventionTitle elements. The title of
invention is should be between two to seven words.
-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes
for the representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventionTitle (#PCDATA)>
<!ATTLIST InventionTitle
    languageCode CDATA #REQUIRED

>
<!--SequenceTotalQuantity
Indicates the total number of sequences in the document.
Its purpose is to be quickly accessible for automatic processing.
-->
<!ELEMENT SequenceTotalQuantity (#PCDATA)>
<!--SequenceData
Data for individual Sequence.
For intentionally skipped sequences see the ST.26 main body document.
-->
<!ELEMENT SequenceData (INSDSeq)>
<!ATTLIST SequenceData
    sequenceIDNumber CDATA #REQUIRED

>
<!--IPOfficeCode
ST.3 code. For example, if the application identification is PCT/IB2013/099999, then IPOfficeCode
value will be "IB" for the International Bureau of WIPO.
-->
<!ELEMENT IPOfficeCode (#PCDATA)>
```

```

<!--ApplicationNumberText
The application identification as provided by the office of filing (e.g. PCT/IB2013/099999)
--&gt;
&lt;!ELEMENT ApplicationNumberText (#PCDATA)&gt;
<!--FilingDate
The date of filing of the patent application for which the sequence listing is submitted in ST.2
format "CCYY-MM-DD", using a 4-digit calendar year, a 2-digit calendar month and a 2-digit day within
the calendar month, e.g., 2015-01-31. For details, please see paragraphs 7 (a) and 11 of WIPO Standard
ST.2.
--&gt;

&lt;!ELEMENT FilingDate (#PCDATA)&gt;
&lt;!--*****
* INSD Part
*****--&gt;</pre>

```

The purpose of the INSD part of this DTD is to define a customized DTD for sequence listings to support the work of IP offices while facilitating the data exchange with the public repositories.

The INSD part is subset of the INSD DTD v1.5 and as such can only be used to generate an XML instance as it will not support the complete INSD structure.

This part is based on:

The International Nucleotide Sequence Database (INSD) collaboration.

INSDSeq provides the elements of a sequence as presented in the GenBank/EMBL/DDBJ-style flatfile formats. Not all elements are used here.

```

-->
<!--INSDSeq
Sequence data. Changed INSD V1.5 DTD elements, INSDSeq_division and INSDSeq_sequence from optional to
mandatory per business requirements.
-->
<!ELEMENT INSDSeq (INSDSeq_length, INSDSeq_moltype, INSDSeq_division, INSDSeq_other-seqids?,
INSDSeq_feature-table?, INSDSeq_sequence)>
<!--INSDSeq_length
The length of the sequence. INSDSeq_length allows only integer.
-->
<!ELEMENT INSDSeq_length (#PCDATA)>
<!--INSDSeq_moltype
Admissible values: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype (#PCDATA)>
<!--INSDSeq_division
Indication that a sequence is related to a patent application. Must be populated with the value PAT.
-->
<!ELEMENT INSDSeq_division (#PCDATA)>
<!--INSDSeq_other-seqids
In the context of data exchange with database providers, the IPOs should populate for each sequence
the element INSDSeq_other-seqids with one INSDSeqid containing a reference to the corresponding
published patent and the sequence identification.
-->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid?)>
<!--INSDSeq_feature-table
Information on the location and roles of various regions within a particular sequence. Whenever the
element INSDSeq_feature-table is used, it must contain at least one feature.
-->
<!ELEMENT INSDSeq_feature-table (INSDFeature+)>
<!--INSDSeq_sequence
```

The residues of the sequence. The sequence must not contain numbers, punctuation or whitespace characters.

-->  
<!ELEMENT INSDSeq\_sequence (#PCDATA)>

<!--INSDSeqid

Intended for the use of IPOs in data exchange only.

Format:

pat|{office code}|{publication number}|{document kind code}|{Sequence identification number}

where office code is the code of the IP office publishing the patent document, publication number is the publication number of the application or patent, document kind code is the letter codes to distinguish patent documents as defined in ST.16 and Sequence identification number is the number of the sequence in that application or patent

Example:

pat|WO|2013999999|A1|123456

This represents the 123456th sequence from WO patent publication No. 2013999999 (A1)

-->

<!ELEMENT INSDSeqid (#PCDATA)>

<!--INSDFeature

Description of one feature.

-->

<!ELEMENT INSDFeature (INSDFeature\_key, INSDFeature\_location, INSDFeature\_quals?)>

<!--INSDFeature\_key

A word or abbreviation indicating a feature.

-->

<!ELEMENT INSDFeature\_key (#PCDATA)>

<!--INSDFeature\_location

Region of the presented sequence which corresponds to the feature.

-->

<!ELEMENT INSDFeature\_location (#PCDATA)>

<!--INSDFeature\_quals

List of qualifiers containing auxiliary information about a feature.

-->

<!ELEMENT INSDFeature\_quals (INSDQualifier+)>

<!--INSDQualifier

Additional information about a feature.

For coding sequences and variants see the ST.26 main body document.

-->

<!--id

Unique identifier for the INSDQualifier to facilitate comparison of versions of a sequence listing specifically having language-dependent qualifier values in different languages.

-->

<!ELEMENT INSDQualifier (INSDQualifier\_name, INSDQualifier\_value?, NonEnglishQualifier\_value?)>

<!ATTLIST INSDQualifier

    id ID #IMPLIED

>

<!--INSDQualifier\_name

Name of the qualifier.

-->

<!ELEMENT INSDQualifier\_name (#PCDATA)>

<!--INSDQualifier\_value

Value of the qualifier. Where the qualifier is language dependent its value must be in the English language and typed with the characters set forth in paragraph 40 (b).

-->

<!ELEMENT INSDQualifier\_value (#PCDATA)>

```
<!--NonEnglishQualifier_value  
Value of a language-dependent qualifier in a language that is not English and typed with the  
characters set forth in paragraph 40 (a). The language is indicated with the attribute  
nonEnglishFreeTextLanguageCode.  
-->  
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[次頁から附属書Ⅲが続く]

### 附属書III

#### 配列表の実例(XML ファイル)

*Version 1.4*

2023年12月8日開催の第11回WIPO標準委員会(CWS)会合で採択

附属書IIIは以下のサイトで閲覧可能。：

[https://www.wipo.int/standards/en/xml\\_material/st26/st26-annex-iii-sequence-listing-specimen.xml](https://www.wipo.int/standards/en/xml_material/st26/st26-annex-iii-sequence-listing-specimen.xml)

[次頁から附属書IVが続く]

## 附属書IV

配列表のXMLインスタンスに用いる  
Unicode 基本ラテンコード表の文字サブセット

*Version 1.3*

2023年12月8日開催の第11回WIPO標準委員会(CWS)会合で採択

アンパサンド文字(&のこと)(0026)は、予め定義された文字実体の一部としてのみ許可される。引用符(0022)、アポストロフィ(0027)、不等号(小号、003C)および不等号(大号、003E)は、それぞれ定義済み実体で表示しなければならない。また、アンパサンド文字(&のこと)(0026)は、属性値や要素の内容の中でアンパサンドとして使用される場合には、予め定義された文字実体で表されなければならない。

Unicode code point	Character	Name
0020		SPACE
0021	!	EXCLAMATION MARK
0022	"	QUOTATION MARK
0023	#	NUMBER SIGN
0024	\$	DOLLAR SIGN
0025	%	PERCENT SIGN
0026	&	AMPERSAND
0027	'	APOSTROPHE
0028	(	LEFT PARENTHESIS
0029	)	RIGHT PARENTHESIS
002A	*	ASTERISK
002B	+	PLUS SIGN
002C	,	COMMA
002D	-	HYPHEN-MINUS
002E	.	FULL STOP
002F	/	SOLIDUS
0030	0	DIGIT ZERO
0031	1	DIGIT ONE
0032	2	DIGIT TWO
0033	3	DIGIT THREE
0034	4	DIGIT FOUR
0035	5	DIGIT FIVE
0036	6	DIGIT SIX
0037	7	DIGIT SEVEN
0038	8	DIGIT EIGHT
0039	9	DIGIT NINE
003A	:	COLON
003B	;	SEMICOLON
003C	<	LESS-THAN-SIGN
003D	=	EQUALS SIGN
003E	>	GREATER-THAN-SIGN
003F	?	QUESTION MARK
0040	@	COMMERCIAL AT
0041	A	LATIN CAPITAL LETTER A
0042	B	LATIN CAPITAL LETTER B
0043	C	LATIN CAPITAL LETTER C
0044	D	LATIN CAPITAL LETTER D
0045	E	LATIN CAPITAL LETTER E
0046	F	LATIN CAPITAL LETTER F
0047	G	LATIN CAPITAL LETTER G
0048	H	LATIN CAPITAL LETTER H
0049	I	LATIN CAPITAL LETTER I
004A	J	LATIN CAPITAL LETTER J

Unicode code point	Character	Name
004B	K	LATIN CAPITAL LETTER K
004C	L	LATIN CAPITAL LETTER L
004D	M	LATIN CAPITAL LETTER M
004E	N	LATIN CAPITAL LETTER N
004F	O	LATIN CAPITAL LETTER O
0050	P	LATIN CAPITAL LETTER P
0051	Q	LATIN CAPITAL LETTER Q
0052	R	LATIN CAPITAL LETTER R
0053	S	LATIN CAPITAL LETTER S
0054	T	LATIN CAPITAL LETTER T
0055	U	LATIN CAPITAL LETTER U
0056	V	LATIN CAPITAL LETTER V
0057	W	LATIN CAPITAL LETTER W
0058	X	LATIN CAPITAL LETTER X
0059	Y	LATIN CAPITAL LETTER Y
005A	Z	LATIN CAPITAL LETTER Z
005B	[	LEFT SQUARE BRACKET
005C	\	REVERSE SOLIDUS
005D	]	RIGHT SQUARE BRACKET
005E	^	CIRCUMFLEX ACCENT
005F	˘	LOW LINE
0060	ˋ	GRAVE ACCENT
0061	a	LATIN SMALL LETTER A
0062	b	LATIN SMALL LETTER B
0063	c	LATIN SMALL LETTER C
0064	d	LATIN SMALL LETTER D
0065	e	LATIN SMALL LETTER E
0066	f	LATIN SMALL LETTER F
0067	g	LATIN SMALL LETTER G
0068	h	LATIN SMALL LETTER H
0069	i	LATIN SMALL LETTER I
006A	j	LATIN SMALL LETTER J
006B	k	LATIN SMALL LETTER K
006C	l	LATIN SMALL LETTER L
006D	m	LATIN SMALL LETTER M
006E	n	LATIN SMALL LETTER N
006F	o	LATIN SMALL LETTER O
0070	p	LATIN SMALL LETTER P
0071	q	LATIN SMALL LETTER Q
0072	r	LATIN SMALL LETTER R
0073	s	LATIN SMALL LETTER S
0074	t	LATIN SMALL LETTER T
0075	u	LATIN SMALL LETTER U
0076	v	LATIN SMALL LETTER V
0077	w	LATIN SMALL LETTER W
0078	x	LATIN SMALL LETTER X
0079	y	LATIN SMALL LETTER Y
007A	z	LATIN SMALL LETTER Z
007B	{	LEFT CURLY BRACKET
007C		VERTICAL LINE
007D	}	RIGHT CURLY BRACKET
007E	~	TILDE