

Technical Working Party on Testing Methods and Techniques**TWM/2/15 Rev.****Second Session****Virtual meeting, April 8 to 11, 2024****Original:** English**Date:** April 9, 2024

WIPO STANDARD ST.26 AND PATENT SEQUENCE DATA*Document prepared by an expert from the World Intellectual Property Organization (WIPO)**Disclaimer: this document does not represent UPOV policies or guidance*

The annex to this document contains a copy of a presentation “WIPO Standard ST.26 and Patent Sequence Data”, to be made by an expert from the World Intellectual Property Organization (WIPO), at the second session of the Technical Working Party on Testing Methods and Techniques (TWM).

[Annex follows]



WIPO Standard ST.26 and Patent Sequence Data
UPOV Technical Working Party on Testing Methods and Techniques
(TWM/2, April 2024)

Emma Francis
IP Data Expert and WIPO Sequence Product
Owner, WIPO

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

WIPO FOR OFFICIAL USE ONLY

1

To be covered today 1

- What is a sequence listing? And how is it used?
- WIPO ST.26 and its benefits for patent applicants/offices
- Introduction to WIPO Sequence
- INSDC and Patentscope databases
- Sequence similarity search tools

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

2

A sequence listing...

2

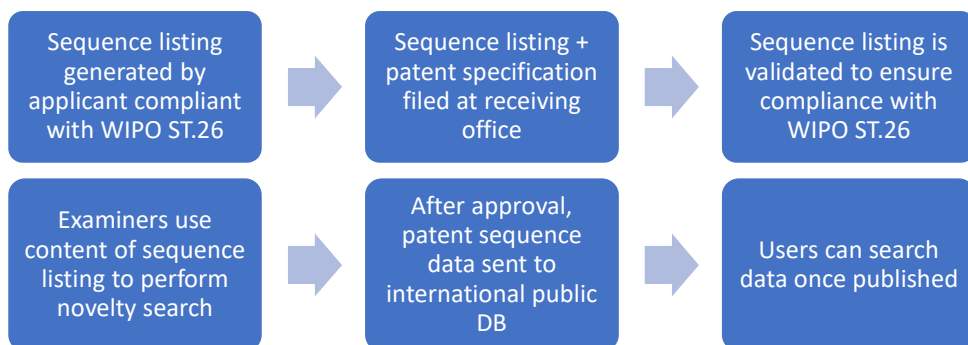
- Contains nucleotide and/or amino acid sequences **disclosed in a patent application** and this forms part of the description.
 - It includes description information about each sequence known as **annotations**
 - Conforms to the requirements of a relevant WIPO Standard (WIPO ST.25 or ST.26)
- Allows for the sequence data of an invention to be searchable:
- Inside an IP Office
 - In publicly available genomic databases



3

How is a sequence listing used (simplified)?

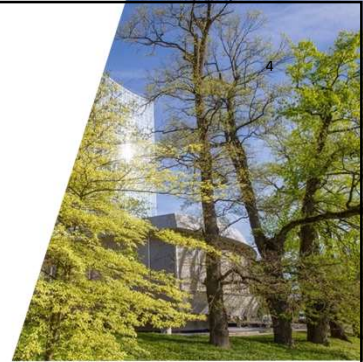
3



4

WIPO Standard ST.26 (1)

- A set of recommendations for the presentation of nucleotide or amino acid information in an XML (eXtensible Markup Language) format
- It is **mandatory** that from July 1, 2022 that **all** sequence listings filed at the international, regional and national levels must be in compliance with this Standard
- Any nucleotide sequences disclosed within a patent specification which include more than 10 specifically defined residues OR for amino acid sequences more than 4 specifically defined amino acids
- Reuses public genomic database vocabularies
- Comprises a single XML document disclosing both sequence data and bibliographic data associating it with the relevant patent application



5

WIPO Standard ST.26 (2)

- Standard comprised of:
 - Main body
 - Annex I: INSDC controlled vocabulary
 - Annex II: ST.26 Document Type Definition (DTD)
 - Annex III: Example sequence listing
 - Annex IV: Allowable character subset
 - Annex V: IP Office data exchange for Offices
 - Annex VI: Guidance document plus sequence examples
 - Annex VII: Recommendations for transformation from WIPO ST.25 to ST.26



6

WIPO ST.26 Benefits (1)

Acceptance of a single sequence listing worldwide

Guidance to ensure agreement on application of sequence rules

Clarifies what sequence disclosures are required or permitted e.g., Annex VI

Enhanced submission quality due to the XML structured format

Increased automation: data validation, analysis and processing



7

WIPO ST.26 Benefits (2)

Data compatibility with data provider requirements

Standardization of feature annotations and locations, qualifiers, sequence variants

Requirement for the inclusion of additional types of sequences: more searchable sequence data



8

WIPO ST.26 example: [Annex III](#)

8

```
<?xml version="1.0" encoding="UTF-8" file="st26-annex-iii-sequence-listing-specimen.xml" software="WIPO Sequence" softwareVersion="1.2.0" productionDate="2022-01-07" originalFreeTextLanguageCode="ja" nonEnglishFreeTextLanguageCode="de">
  <ApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
    <FilingDate>2015-01-31</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
    <FilingDate>2014-01-30</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="ja">出願者株式会社</ApplicantName>
  <ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantNameLatin>
  <InventorName languageCode="ja">特許 太郎</InventorName>
  <InventorNameLatin>Taro Tokkyo</InventorNameLatin>
  <InventionTitle languageCode="ja">efgタンパク質をロードするマウスabcd-1遺伝子</InventionTitle>
  <InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
  <SequenceTotalQuantity>1</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1">
    <INSDSeq>
      <INSDSeq_length>133</INSDSeq_length>
      <INSDSeq_multype>DNA</INSDSeq_multype>
      <INSDSeq_division>PAT</INSDSeq_division>
      <INSDSeq_feature-table>
        <INSDFeature>
          <INSDFeature_key>source</INSDFeature_key>
          <INSDFeature_location>1..133</INSDFeature_location>
          <INSDFeature_qual>
            <INSDQualifier>
              <INSDQualifier_name>organism</INSDQualifier_name>
              <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
              <INSDQualifier_name>mol_type</INSDQualifier_name>
              <INSDQualifier_value>genomic DNA</INSDQualifier_value>
            </INSDQualifier>
          </INSDFeature_qual>
        </INSDFeature>
      </INSDSeq_feature-table>
      <INSDSeq_sequence>atgaaataaaacataaaaggatgataaaatgagattgatataaaaaaggttttagagttagcagagaaggattttgagacgcatggagagagacaagggcattaataaaggataaacatttgacaata</INSDSeq_sequence>
    </INSDSeq>
  </SequenceData>
</?xml>
```



9

WIPO Sequence: introduction

9

- **Free** desktop tool developed by WIPO to support the authoring, validation and generation of sequence listings compliant with WIPO Standard ST.26
- All installation packages and user support documentation can be downloaded from the WIPO Sequence homepage at: <https://www.wipo.int/standards/en/sequence>
- Interface is provided in the 10 Patent Cooperation Treaty (PCT) publication languages and for 3 operating systems
- Use of WIPO Sequence simplifies ST.26 XML creation with a user-friendly interface: no need to ever directly edit an XML file
- First stable release, version 1.0.0 published in November 2020 for user testing
- Current version in production is version 2.3.0
- Note: all generated sequence listings and project data entered into the tool are stored locally (on the users computer only).



10

INSDC databases

10

International Nucleotide Sequence Data Collaboration (INSDC)

Collaboration between:

- DDBJ: DNA Databank of Japan
- EMBL-EBI: The European Bioinformatics Institute
- NCBI: National Center for Biotechnology Information (GenBank)

INSDC databases are publicly searchable and there is no restrictions on the access or use of this data

IP Offices that submit published/issued application sequence data to INSDC include:

- European Patent Office
- Japanese Patent Office
- Korean Intellectual Property Office
- United States Patent and Trademark Office

Data which is ingested must comply with strict guidelines including compliance with the controlled vocabulary and DTD

WIPO ST.26 Annex I is a complete list of this controlled vocabulary



11

PATENTSCOPE

11

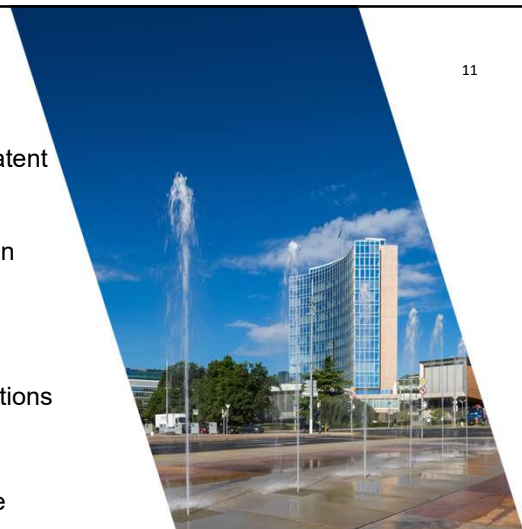
PATENTSCOPE is a free database containing 115 million patent documents which is maintained by WIPO including:

- All PCT applications which are Open for Public Inspection (OPI)
- National collections made available to WIPO
- Non Patent Literature (NPL) data including [IEEE](#) publications

<https://patentscope.wipo.int/search/en/search.jsf>

PATENTSCOPE also publishes on a weekly basis sequence listings which are received but only as ZIP files at:

<https://patentscope.wipo.int/search/en/sequences.jsf>



12

Similarity Search Examples - INSDC

12

- National Center for Biotechnology Information (NCBI) offers [BLAST free comparison tool](#) which is also used by European Nucleotide Archive (ENA)
- “The Basic Local Alignment Search Tool (BLAST) find regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches”



13

Soybean example in BLAST

13

Job Title: Tropical soybean search
 RID: 0BE7YDJ016 Search expires on 03-28 23:19 pm Download All
 Program: BLASTN Citation
 Database: Genome (Glycine_max_v4.0 reference, Annotation Release 104) See details
 Query ID: lclQuery_2835629
 Description: None
 Molecule type: dna
 Query Length: 1001
 Other reports: Distance tree of results MSA viewer

Filter Results
 Organism: only top 20 will appear exclude
 Type common name, binomial, taxid or group name
 + Add organism
 Percent Identity: [] to [] E value: [] to [] Query Coverage: [] to []
 Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

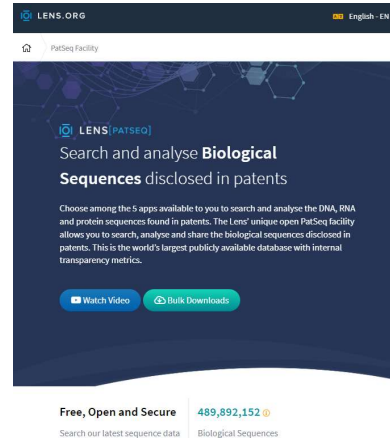
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Glycine max cultivar Williams 82 chromosome 1_Glycine_max_v4.0	Glycine max	1834	4481	100%	0.0	99.70%	57932355	NC_016088.4
Glycine max cultivar Williams 82 unplaced genomic scaffold_Glycine_max_v4.0 scaffold_117	Glycine max	693	3655	99%	0.0	97.77%	64814	NW_024464762.1
Glycine max cultivar Williams 82 chromosome 15_Glycine_max_v4.0	Glycine max	688	3344	100%	0.0	93.70%	53754295	NC_038251.2
Glycine max cultivar Williams 82 chromosome 9_Glycine_max_v4.0	Glycine max	671	4811	99%	0.0	96.77%	47227184	NC_038244.2
Glycine max cultivar Williams 82 chromosome 3_Glycine_max_v4.0	Glycine max	459	1940	99%	6e-127	93.77%	46951866	NC_016090.4
Glycine max cultivar Williams 82 chromosome 11_Glycine_max_v4.0	Glycine max	440	1470	86%	1e-123	93.14%	39643745	NC_038247.2
Glycine max cultivar Williams 82 chromosome 10_Glycine_max_v4.0	Glycine max	425	1487	91%	6e-117	93.17%	51638687	NC_038246.2
Glycine max cultivar Williams 82 chromosome 6_Glycine_max_v4.0	Glycine max	424	3226	91%	2e-116	93.17%	50945864	NC_038242.2
Glycine max cultivar Williams 82 chromosome 13_Glycine_max_v4.0	Glycine max	399	1022	62%	4e-109	97.06%	45225048	NC_038249.2



14

Similarity Search Examples – Patent Offices

- [PatSeq Finder](#): search based on BLAST for sequence data saved in The Lens patent database
- “The Lens’ unique open PatSeq facility allows you to search, analyze and share biological sequences disclosed in patents. This is the world’s largest publicly available database with internal transparency metrics.”
- Requires an account for use
- Includes sequence listings in original WIPO ST.26 format



15

Soybean example in PatSeq Finder

14

PatSeq Home Data Text Explorer Finder Analyzer

Results

Sequence Search (BLASTn 1001bp vs patseq-n) Start another search

Query: agggacctgtgtcagtggtataaaactcagactggacc... (Length: 1001bp)

Hide summary of result average and medians

Number of unique patent documents	34
Average number of HSP per hit	1.1 (max 2)
Average hit sequence length	637 (min 85 - max 2,048)
Average alignment length (all HSPs)	131 (min 34 - max 426)
E value range	avg: 1.35e-6 (min 0 - max 1.99e-5)

Location in document Grants, in claims (0) Grants (0) Applications, in claims (75) Applications (136)

Showing 1 to 25 of 136 filtered hits - 157 total hits

Sequence	Coverage	Similarity	Alignment length	E-value	BLAST score	Patent number
SEQ ID 137,421 US 20040031072 A1 Sequence length: 1,013bp	41.3%	98.8%	413bp	0	737.93 bits	US 20040031072 A1

Soy nucleic acid molecules and other molecules associated with transcription plants and uses thereof for plant improvement

US, Published Feb 12, 2004, Filed Apr 26, 2003
Applicants: LA ROSA THOMAS J., ZHOU YIHUA, KOVALIC DAVID K., CAO YONGWEI
Organism: Glycine max

Status: abandoned

Abstract: Polynucleotides useful for improvement of plants are provided. In particular, polynucleotide sequences are provided from plant sources. Polypeptides encoded by the polynucleotide sequences are also provided. The disclosed polynucleotides and polypeptides find use in production of transgenic plants to produce plants having improved properties.

16

Any questions:
standards@wipo.int

WIPO Standard ST.26:
<https://www.wipo.int/export/sites/www/standards/en/pdf/03-26-01.pdf>

WIPO Standards webinars:
https://www.wipo.int/meetings/en/topic.jsp?group_id=330

© WIPO, 2024



Attribution 3.0 IGO
(CC BY 3.0 IGO)

The CC license does not apply to non-WIPO content in this presentation.

Photo credits:

