

STN[®]

Coverage of Asian authorities in patent
sequence databases

Robert Austin – FIZ Karlsruhe

Agenda

2

- Public sequence databases
- STN[®] sequence databases
- Comparison of Asian content (JP, KR, CN)
- Results of two search examples
- GenomeQuest GQ-PAT coverage
- Conclusions

There are three web resources that provide searchable Asian patent sequence data

- National Center for Biotechnology Information (NCBI) of the U.S. National Library of Medicine
 - www.ncbi.nlm.nih.gov
- DNA DataBank of Japan (DDBJ)
 - www.ddbj.nig.ac.jp
- European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI)
 - www.ebi.ac.uk
- SwissProt, PIR, etc do not cover patents

The EPO, USPTO and JPO/KIPO rely on the NCBI, EBI and DDBJ, respectively, to provide an interface for searching patent sequence data.

The International Nucleotide Sequence Database Collaboration (INSDC)

4

The screenshot shows the INSDC website with a navigation bar containing 'ABOUT INSDC', 'POLICY', and 'ADVISORY BOARD'. A blue box highlights the URL 'The INSDC: <http://www.insdc.org/>'. The main content area is titled 'International Nucleotide Sequence Database Collaboration' and contains a bulleted list of facts. A blue box on the left lists the logos for NCBI, DDBJ, and EMBL. Two blue boxes on the right contain explanatory text: one states that the collaboration is for nucleotide sequences only, and the other notes that Asian patent nucleotide sequence data (JP, KR) is provided to DDBJ by JPO and KIPO respectively. The footer of the website displays the INSDC logo and name.

ABOUT INSDC POLICY ADVISORY BOARD

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

International Nucleotide Sequence Database Collaboration

- The International Nucleotide Sequence Databases (INSDB) have been developed and maintained collaboratively between [DDBJ](#), [EMBL](#), and [GenBank](#) for over 18 years.
- The INSDC advisory board, the [International Advisory Committee](#), consists of representatives of each of the databases' advisory bodies. At their most recent meeting, the committee unanimously endorsed and reaffirmed the existing policy of mutual exchange of data among the three databases that make up the INSDC, which is stated below.
- Individuals submitting data to the international sequence databases should be aware of the [INSDC policy](#).

How to submit data

- For full details of how to submit data to the databases, please contact your preferred database partner.
- [DDBJ](#), [EMBL](#), [GenBank](#)
- The INSDC Feature Table Definition Document is available [here](#).

NCBI

DDBJ

EMBL
Nucleotide
Sequence
Database









The INSDC collaboration and mutual exchange of data (i.e. Genbank), is for nucleotide sequences only.

Asian patent nucleotide sequence data (JP, KR) is provided to DDBJ by JPO and KIPO respectively.

INSDC
International Nucleotide Sequence Database Collaboration

In addition, EMBL-EBI provides access to Asian patent peptide sequence data

5

	Nucleotide		Peptide	
	JP	KR	JP	KR
EMBL				
NCBI				
DDBJ				

EMBL-EBI provides several different patent sequence data collections for searching

The screenshot shows the 'Available Databanks' page on the EMBL-EBI website. The page is organized into several sections with expandable/collapsible options. Annotations include:

- A blue box at the top right contains the text: "Asian (JP, KR) patent sequences are included in the EMBL Nucleotide Database and the Patent Proteins collection." Blue lines connect this box to the 'EMBL' checkbox and the 'Patent Proteins' checkbox.
- A red box at the bottom right contains the text: "The EPO nucleotide sequence database, and the EPO, JPO, KIPO and USPTO protein databases may also be searched separately." Red lines connect this box to the 'Patent DNA' checkbox, the 'EPO Proteins' checkbox, the 'JPO Proteins' checkbox, and the 'USPTO Proteins' checkbox.
- A blue box on the right side contains the text: "EMBL-EBI SRS: <http://srs.ebi.ac.uk/>"

The interface includes the following sections and items:

- Available Databanks** (Expand all / Collapse all)
- Literature, Bibliography and Reference Databases**
- Gene Dictionaries and Ontologies**
- Nucleotide sequence databases**
 - EMBL
 - EMBL (Contigs expanded)
 - EMBL ID/Accession Mapping
 - IMGT/HLA
 - GR Genes
 - LiveLists
 - Patent DNA
 - EMBL (Annotated Cons)
 - EMBL MGA
 - IPD-KIR
 - GR Transcripts
 - EMBL (Contig)
 - EMBL (Coding Sequences)
 - IMGT/LIGM-DB
 - Genome Reviews
 - RefSeq Genome
- Nucleotide sequence databases - subsections**
 - EMBL (Updates)
 - EMBL (Release)
 - EMBL (Whole Genome Shotgun)
 - EMBL (Whole Genome Shotgun release)
 - EMBL (Whole Genome Shotgun updates)
 - EMBL (Contig release)
 - EMBL (Contig updates)
 - EMBL (Contigs expanded release)
 - EMBL (Contigs expanded update)
 - EMBL (Annotated Cons release)
 - EMBL (Annotated Cons updates)
 - EMBL (Release, Deleted)
 - EMBL (Whole Genome Shotgun Master)
 - RefSeq Genome (Updates)
- Nucleotide related databases**
- UniProt Universal Protein Resource**
- Other protein sequence databases**
 - Active protein sequence databases**
 - Patent Proteins
 - EPO Proteins
 - JPO Proteins
 - USPTO Proteins
 - KIPO Proteins
 - IPI
 - IPI History
 - MHCBN

EMBL-EBI patent sequence records provide minimal bibliographic and text data

General Information			
Primary Accession #	DL078725		
Accession #	DL078725		
SRS Entry ID	EMBL:DL078725		
Molecule Type	linear unassigned DNA		
Sequence Length	45		
Entry Division	SYN (<i>Synthetic</i>)		
Entry Data Class	PAT (<i>Patent</i>)		
Sequence Version	DL078725.1		
Creation Date	25-AUG-2008		
Modification Date	25-AUG-2008		
EMBL-SVA	DL078725		
Description			
Description	Methods and Compositions for the Inhibition of Gene Exp		
Keywords	JP 2008500838-A/1350.;		
Organism	synthetic construct		
Organism Classification	other sequences; artificial sequences.		
References			
1.	Olson,D.; Sooch,M.P.; Goodwin,N.; Sheikhhnejad,G.; Methods and Compositions for the Inhibition of Gene Expression Patent number JP2008500838-A /1350, 17-JAN-2008. Gholamreza Sheikhhnejad,Mina Patel Sooch,Neal Goodwin,David Olson.		
	Patent	JP2008500838 Esp@cenet CiteXplore	
	Position	1-45	
Features			
Key	Location	Qualifier	Value
source	1..45	organism	synthetic construct
		mol_type	unassigned DNA
		db_xref	taxon:32630
Sequence			
Characteristics	Length: 45 BP, A Count:9, C Count:16, G Count:19, T Count:1, Others Count:0		
Sequence	<pre>>embl DL078725 DL078725 Methods and Compositions for the Inhibition of G. CCAGAGCCTGGCGGCGCGAGCCGGGCACCCGGGCGCGAAGAAAAGG</pre>		

EMBL JP/KR nucleotide sequence records, like this one, are typically also available at NCBI and DDBJ.

JP/KR sequences are the slowest to update, e.g. this JP record was created 7 months after publication.

Korean Intellectual Property Office (KIPO) sequence data were added in Q1 2008

8

The March 26th, 2008 (i.e. latest) release of the KIPO protein database has 113,555 records.

Display Name **KIPO Proteins**

Name KIPO_PRT

Status The current release has 113555 entries and was indexed 26-Mar-2008.

Description Protein sequences extracted from patent applications to the [Korean Intellectual Property Office](#) (KIPO).

WWW <http://www.ebi.ac.uk/>

Ftp <ftp://ftp.ebi.ac.uk/pub/databases/embl/patent/>

Data-fields in SRS	Field Name	Short Name	Type	No. of Keys	No. of Entry References	Indexing Date	Status
	AllText	all	group	0	0		not indexed
	AllIDs	allids	group	0	0		not indexed
	ID	id	id	113555	113555	26-Mar-2008	ok
	Topology	topo	index	0	0	26-Mar-2008	ok
	Molecule	mol	index	1	113555	26-Mar-2008	ok
	Data Class	cla	index	1	113555	26-Mar-2008	ok
	Division	div	index	0	0	26-Mar-2008	ok
	Sequence Length	sl	num	1520	113555	26-Mar-2008	ok
	AccNumber	acc	index	113555	113555	26-Mar-2008	ok
	Primary Accession Number	pac	index	113555	113555	26-Mar-2008	ok
	Sequence Version	sv	index	0	0	26-Mar-2008	ok
	Entry Creation Date	crd	date	1	113555	26-Mar-2008	ok
	Entry Creation Release	relc	num	0	0	26-Mar-2008	ok
	Last Update Date	crlu	date	0	0	26-Mar-2008	ok

Tip: unlike other patent authorities, the KIPO records use an *application number and date*

Application number KR 1019900014814 was filed on September 19th, 1990. The STN search format for this number is:
=> S KR1990-14814/AP

General Information			
Primary Accession #	DI512332		
Accession #	DI512332		
SRS Entry ID	KIPO_PRT:DI512332		
Molecule Type	PRT		
Sequence Length	273		
Entry Data Class	STANDARD		
Creation Date	21-FEB-2008		
UniParc	UPI000002D110		
Description			
Description	Vaccine against lyme disease.		
Keywords	KR 1019900014814-A/2.;		
Organism	Borrelia burgdorferi		
Organism Classification	Borrelia burgdorferi Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.		
References			
1.	Simon; M.M.; Schaible; U.E.; Eichmann; K.; Kramer; M. and Reinhard; W.; Vaccine against lyme disease Patent: KR 1019900014814-A/2 19-SEP-1990.		
	Position	1-273	
Features			
Key	Location	Qualifier	Value
source	1..273		
Sequence			
Characteristics	Length: 273 AA		
Sequence	<pre>>kipo_prt DI512332 DI512332 Vaccine against lyme disease. MKKYLLGIGLILALIACKQNVSSLDEKNSVSVLDLPGEMNVLVSKEKNKDGKYDLIATVDK</pre>		

- Sequences from the *basic* patents of the 41 authorities of the Derwent World Patents Index®
 - Including from Asia: JP, CN, KR, TW, IN, SG
- Largest value-added patent sequence database
- Used routinely by all major patent offices*
- Bibliography, enhanced title, abstract, indexing and patent location provided for each sequence
- Organism name, Sequence Length, Molecule Type and enhanced features/annotations
- Updated every two weeks
- 1981 - present

* See page 11: http://www.trilateral.net/projects/biotechnology/search_guidebook_vers_1.pdf

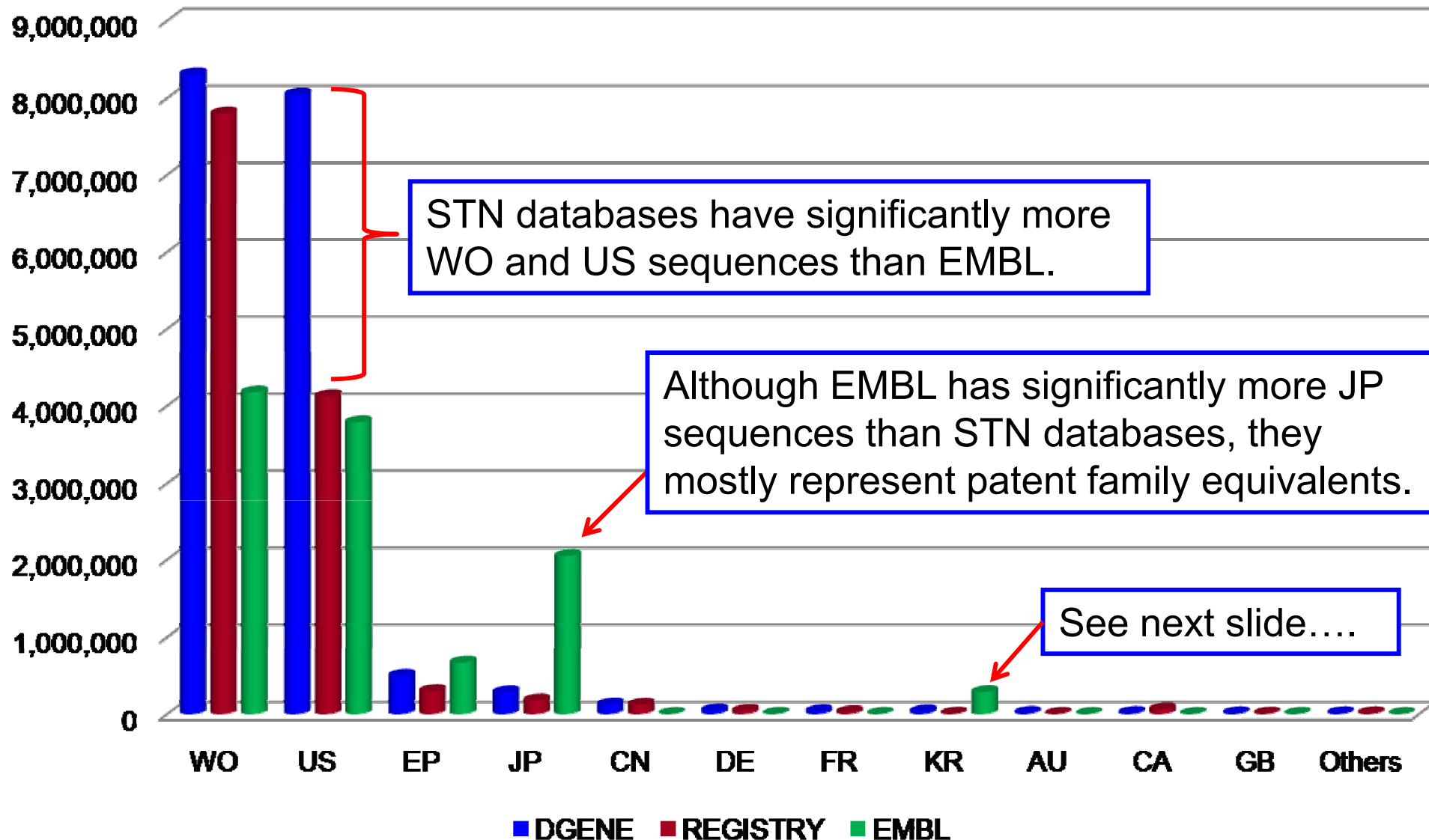
CAS REGISTRY

11

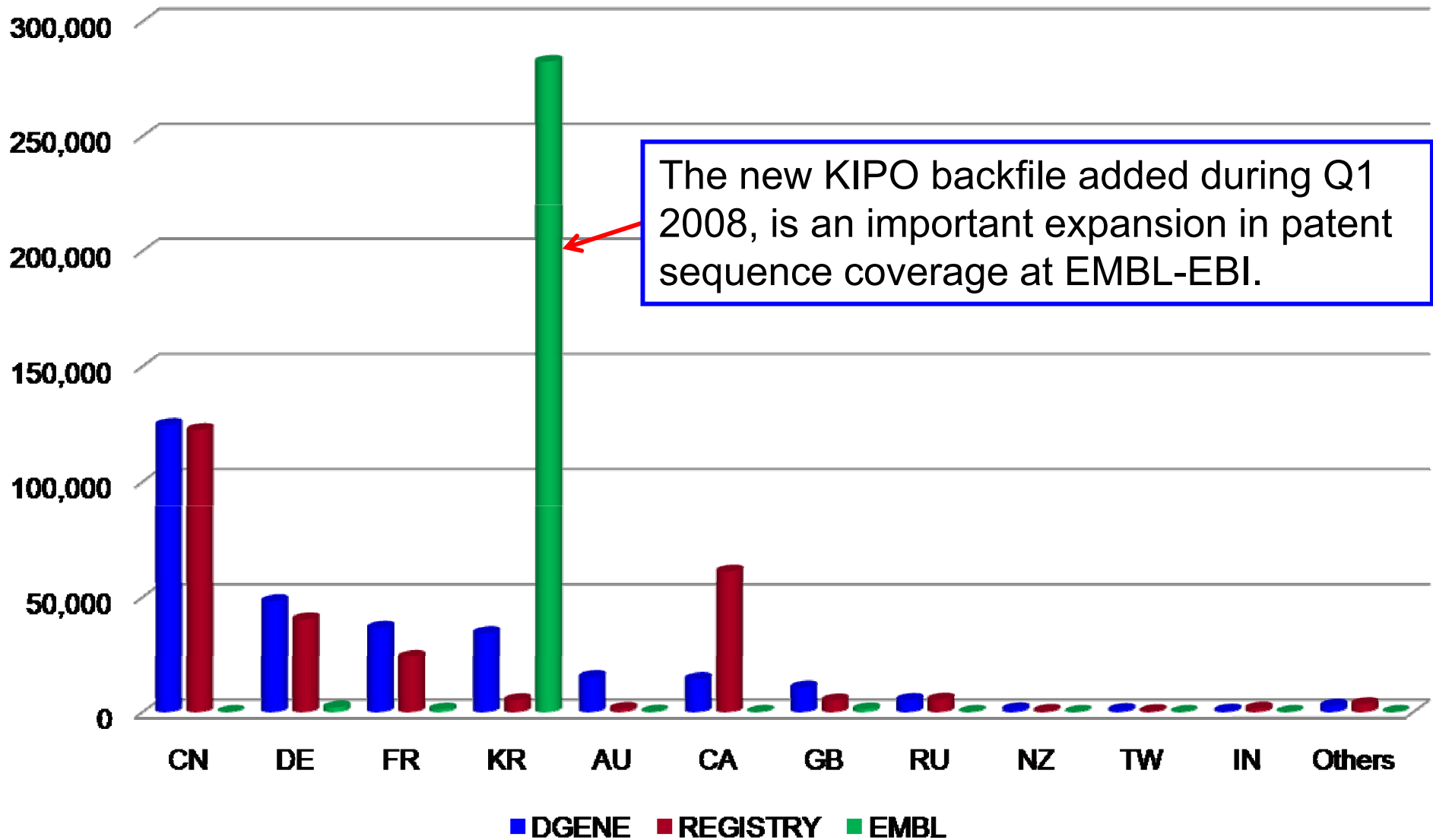
- Sequences from *basic* patents of the 57 patent-issuing authorities in CaplusSM
 - Including from Asia: JP, CN, KR, TW, IN, SG, EG, HK
- Sequences from >3000 life science journals covered in CAplus
- Publication number, patent location, and standardized nomenclature for each sequence
- Organism name, sequence length, molecule type, SEQ ID number and features/annotations
- Updated daily
- 1907 - present

Sequence record counts by patent authority in DGENE, REGISTRY and EMBL

12



Sequence record counts by patent authority in DGENE, REGISTRY and EMBL (cont.)



Approximate timeliness and backfile coverage of Asian authorities in patent sequence databases

	JP		KR		CN	
	Backfile	Latest	Backfile	Latest	Backfile	Latest
EMBL	1979	Jul 2008	~1990	Jun 2007	X	X
DGENE (Basics)	1981	Dec 2008	1997	Aug 2008	1993	Nov 2008
REG (Basics)	1916	Jan 2009	~1996	Jan 2009	~1996	Dec 2008

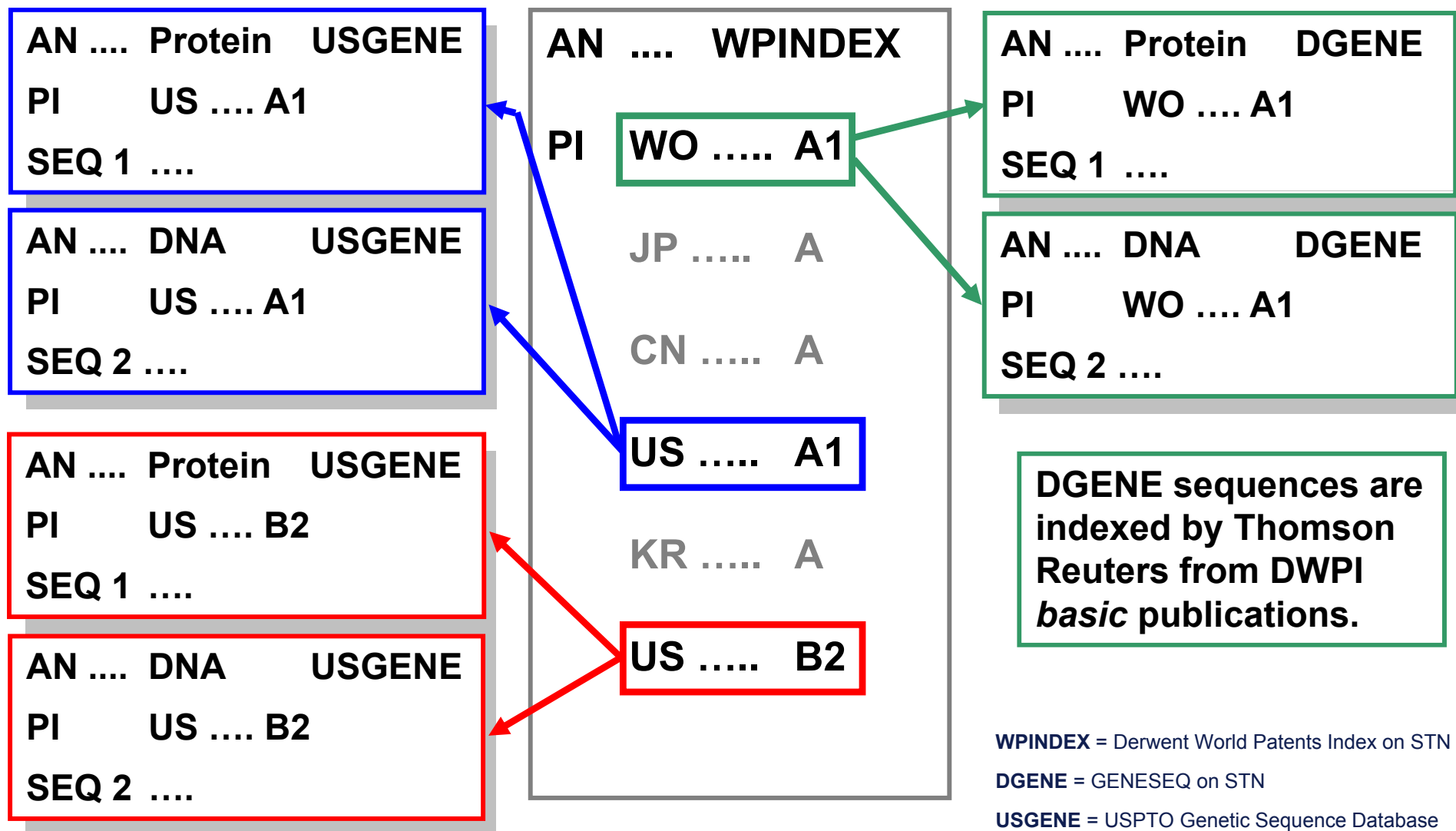
(Status as of February 6th, 2009. Note: EMBL-EBI does not cover sequences from Chinese Patents.)

SequenceBase USGENE

15

- Sequences from all relevant USPTO published patent applications and granted (issued) patents
- Bibliography, title, abstract, claims and patent location provided for each sequence
- Organism name, sequence length, molecule type, SEQ ID number and features/annotations
- Updated weekly – within 3 days of publication
- May provide additional access to Asian patents via patent family databases such as DWPISM
- 1982 – present

USGENE and DGENE may find relevant Asian patents via non-Asian family members



A BLAST® search example shows the importance of using DGENE and USGENE for complete Asian patent retrieval

Search Question:

Find all patent references to Interferon β -1a, or other very similar proteins (i.e. 95% or higher).

```
1 MSYNLLGFLQ RSSNFQCQKL LWQLNGRLEY CLKDRMNFDI PEEIKQLQQF
51 QKEDAALTIY EMLQNIFAIF RQDSSSTGWN ETIVENLLAN VYHQINHLKT
101 VLEEKLEKED FTRGKLMSSL HLKRYYGRIL HYLKAKEYSH CAWTIVRVEI
151 LRNFYFINRL TGYLRN
```

(Search conducted on February 6th, 2009)

The best answer from DGENE...

```
L4 ANSWER 1 OF 3301 DGENE COPYRIGHT 2009 THOMSON
AN ATT38374 protein DGENE
TI New polypeptide useful for treating e.g. Crohn's disease, colorectal
sarcoma, allergic rhinitis, Addison's disease, asthma, rheumatoid
arthritis, ulcerative colitis comprises single chain fragment
crystallizable regions of an immunoglobulin.
IN Farrington G K; Saeed-Kothe A; Garber E; Lugovskoy A A
PA (BIOJ) BIOGEN IDEC MA INC.
PI WO 2008143954 A2 20081127
AI WO 2008-US6260 20080514
PRAI US 2007-930227P 20070514
DT Patent
LA English
OS 2009-A32547 [01]
CR N-PSDB: ATT38375
DESC Human antibody construct pEAG2149 protein SEQ ID: 39.
PSL Example 11; SEQ ID NO 39
SCORE 341 100% of query self score 341
```

3,3301 sequence hits
were found in DGENE.

All answers retrieved in a DGENE
sequence search come from a
DWPI *basic* publication – most often
a WIPO/PCT published application.

BLASTALIGN

```
Query = 166 letters
Length = 629
Score = 341 bits (874), Expect = 1e-98
Identities = 166/166 (100%), Positives = 166/166 (100%)
Query: 1 MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIY
MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIY
Sbjct: 1 MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIY
. . . .
```

The best answer from USGENE...

L5 ANSWER 1 OF 7209 USGENE COPYRIGHT 2009

AN 20090030183.2 Protein USGENE

TI Interferon Beta-Like Molecules (PublishedApplication)

IN Rasmussen Poul Baad (Soeberg, DK); Drustrup Joern (Farum, DK);
Rasmussen Grethe (Farum, DK); Pedersen Anders Hjelholt (Lyngby, DK);
Schambye Hans Thalsgard (Frederiksberg C., DK); Andersen Kim Vilbour
(Broenshoej, DK); Bornaes Claus (Hellerup, DK)

PA Maxygen ApS; Maxygen Holdings Ltd

PI US 20090030183 A1 20090129

AI US 2008-206275 20080908

PSL Claim 88; SEQ ID NO 2

DT Patent

SCORE 341 100% of query self score 341

BLASTALIGN

Query = 166 letters

Length = 166

Score = 341 bits (874), Expect = 4e-99

Identities = 166/166 (100%), Positives = 166/166 (100%)

Query: 1 MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIY
MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIY

Sbjct: 1 MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIY

Query: 61 EMLQNIFAIFRQDSSSTGWNENLLANVYHQINHLKTVLEEKLEKEDFTRGKLMSSL
EMLQNIFAIFRQDSSSTGWNENLLANVYHQINHLKTVLEEKLEKEDFTRGKLMSSL

Sbjct: 61 EMLQNIFAIFRQDSSSTGWNENLLANVYHQINHLKTVLEEKLEKEDFTRGKLMSSL

. . . .

7,209 sequence hits were found in USGENE.

USGENE covers both U.S. published applications and granted (issued) patents.

DGENE and USGENE BLAST search results for *Interferon β -1a* – Asian retrieval is in (red)

	Sequences > 95% (CN+JP+KR)	Publication Numbers (CN+JP+KR)	DWPI records (CN+JP+KR)
DGENE	3,301 (8)	154 (7)	154 (74)
USGENE	7,209 (0)	259 (0)	107 (55)
Overlap	-	37 (0)	92 (47)
Total Unique	-	376 (7)	169 (82)

Important conclusion: both DGENE and USGENE are essential for complete Asian patent retrieval via an intermediary patent family database, e.g. DWPI.

An example of a DWPI record uniquely retrieved by USGENE with Asian patent family members

```

AN 2002-750669 [81] WPINDEX
TI New stabilized human serum albumin-free pharmaceutical compositions of
DC B04
IN BABUKA S; CHEN B; CHOE M; HORA M; SHIRLEY B A; TELLERS M
PA (CHIR-C) CHIRON CORP; (NOVS-C) NOVARTIS VACCINES & DIAGNOSTICS INC ...
PI WO 2002080976 A2 20021017 (200281)* EN 71[23]
  US 20020172661 A1 20021121 (200301) EN <--
  NO 2003004492 A 20031125 (200407) NO
  EP 1381432 A2 20040121 (200410) EN
  KR 2003097825 A 20031231 (200427) KO
  AU 2002241769 A1 20021021 (200433) EN
  JP 2004529917 W 20040930 (200465) JA 112
  US 20040191219 A1 20040930 (200465) EN <--
  CN 1511053 A 20040707 (200467) ZH
  CZ 2003002735 A3 20050316 (200522) CS
  US 6887462 B2 20050503 (200530) EN
  US 20050142110 A1 20050630 (200543) EN
  AU 2002241769 B2 20060601 (200705) EN
  CN 1313148 C 20070502 (200758) ZH
  HU 2003003732 A1 20070928 (200780) HU
  US 7371373 B2 20080513 (200834) EN
  US 7399463 B2 20080715 (200848) EN <--
  US 20080193415 A1 20080814 (200856) EN <--
  JP 2008285499 A 20081127 (200880) JA 36
  . . .
  
```

Highlighted in red are the U.S. family members retrieved by USGENE.

This example has JP, CN and KR patent family members, including a patent granted in China.

A BLAST® search example shows that also including REGISTRY and NCBI is critical for a complete Asian patent sequence search

Search Question:

Find all patent references to *Pseudomonas poae** 16S ribosomal RNA gene (85% or higher match).

* *Pseudomonas poae* is a fluorescent, Gram-negative bacterium isolated from the phyllosphere of grasses and, along with other similar bacteria, has potential utility in biofertilizers.

(Search conducted on February 5th, 2009)

Pseudomonas poae 16S ribosomal RNA gene BLAST search results – Asian retrieval is in (red)

	Sequences > 85% (CN+JP+KR)	Publication Numbers (CN+JP+KR)	DWPI records (CN+JP+KR)
REG	41 (22)	34 (15)	32 (26)
NCBI	61 (33)	61 (33)	22 (20)
DGENE	34 (16)	32 (15)	32 (30)
USGENE	48 (0)	51 (0)	19 (17)
Total Unique	-	125 (50)	49 (40)

Note: all four patent sequence databases found unique Asian patent sequence hits, either directly or indirectly via DWPI patent family member relationships.

An example DWPI record uniquely retrieved by the DGENE BLAST search

24

```
AN 2009-A87653 [04] WPIX
DNC C2009-036829 [04]
TI Pseudomonas species strain for microorganism fertilizer, is formed
such that it resolves water insoluble phosphate fixed to the soil
DC B04; D16
IN HANG-YEON W; HYUNG-JUN N; JANG-SI
PA (RURA-N) RURAL DEV ADMINISTRATION
CYC 1
PIA KR 2008058522 A 20080626 (200904)* KO 7[4] <--
ADT KR 2008058522 A KR 2006-132269 20061222
PRAI KR 2006-132269 20061222
AB KR 2008058522 A UPAB: 20090116
NOVELTY - The Pseudomonas sp CL-1 (Culture collection KACC 91282P)
strain is formed such that it resolves water insoluble phosphate
fixed to the soil.
DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included
for a method for performing usability of the water insoluble
phosphate, which involves using the microorganism fertilizer.
USE - Pseudomonas sp CL-1 (Culture collection KACC 91282P)
strain for microorganism fertilizer (Claimed).
ADVANTAGE - The Pseudomonas sp CL-1 (Culture collection KACC
91282P) strain enables to prevent soil environment, and ensures
saving phosphatic chemical fertilizer.
DESCRIPTION OF DRAWINGS - The drawing shows a phase difference
microphotograph of the pseudomonas SP.
```

This Korean document is too recent to be covered by NCBI/EMBL/DDBJ.

An example DWPI record uniquely retrieved by the USGENE BLAST search

AN 2003-665372 [63] WPIX
DNC C2003-180860 [63]
DNN N2003-531089 [63]
TI New polyhydroxyalkanoate used as charge control agent in, e.g. toner binder
DC A23; A89; D16; G08; P84; S06
IN FUKUI S; FUKUI T; HONMA T; IMAMURA T; KENMOKU T; NOMOTO T;
PA (CANO-C) CANON KK; (FUKU-I) FUKUI T; (HONM-I) HONMA T; (IMAM-I) IMAMURA T; (KENM-I) KENMOKU T; (NOMO-I) NOMOTO T
PIA EP 1236755 A2 20020904 (200363)* EN 133
KR 2002070871 A 20020911 (200363) KO
US 20030013841 A1 20030116 (200363) EN
JP 2003306534 A 20031031 (200374) JA 89
JP 3592306 B2 20041124 (200477) JA 134
US 20050250191 A1 20051110 (200574) EN
US 7045321 B2 20060516 (200633) EN
KR 522483 B 20051018 (200680) KO
US 7408017 B2 20080805 (200857) EN
. . . .
TECH BIOTECHNOLOGY - Preferred Components: The culture medium includes, aside from compounds (14), a polypeptone or a yeast extract.
Preferred Microorganism: The microorganism belongs to genus Pseudomonas. Preferably, the microorganism is Pseudomonas cichorii YN2 (FERM BP-7375), Pseudomonas cichorii H45 (FERM BP-7374), Pseudomonas jessenii P161 (FERM BP-7376), and/or Pseudomonas putida P91 (FERM BP-7373).

Highlighted in red are the U.S. family members retrieved by USGENE.

This example has JP, CN and KR patent family members, including a patent granted in Japan.

An example DWPI record uniquely retrieved by the REGISTRY BLAST search

26

```
AN 2008-K98417 [65] WPIX
DNC C2008-318651 [65]
TI Application of Pseudomonas (Pseudomonas sp.) BC001 CGMCC NO.2223 on
biological removal of nitrogen-containing heterocyclic compounds
DC D15; D16
IN BAI Y; SUN Q; TANG X; WEN D;
PA (UYBE-N) UNIV BEIJING
CYC 1
PIA CN 101195812 A 20080611 (200865)* ZH 16[0] <--
ADT CN 101195812 A CN 2007-10303932 20071221
PRAI CN 2007-10303932 20071221
AB CN 101195812 A UPAB: 20081013
```

Reminder: NCBI/EMBL/DDBJ do not provide sequences from Chinese patents.

NOVELTY - Application of Pseudomonas (Pseudomonas sp.) BC001 CGMCC NO.2223 on biological removal of nitrogen-containing heterocyclic compounds.

USE - Application of pseudomonas (pseudomonas sp.) BC001 CGMCC NO.2223 on biological removal of nitrogen-containing heterocyclic compounds.

ADVANTAGE - The BCO01 bacterium has no resistance for kanamycin sulfate, chloramphenicol, tetracycline and spectinomycin, ensuring ecological safety during use.

An example DWPI record uniquely retrieved by the NCBI BLAST search

27

```
AN 2006-666212 [69] WPIX
DNC C2006-203802 [69]
TI Microorganism capable of degrading dioxins Pseudomonas sp. Ph-03
   which economically and efficiently purifies and recovers environment
DC D16; E13; J01
IN CHANG Y S; HONG H B; KIM H I; NAM I H;
PA (PROB-N) PROBIONIC INC
CYC 1
PIA KR 2005101388 A 20051024 (200669)* K
   KR 607031 B1 20060801 (200714) K
ADT KR 2005101388 A KR 2004-26597 20040419;
   KR 607031 B1 KR 2004-26597 20040419
FDT KR 607031 B1 Previous Publ KR 2005101388 A
PRAI KR 2004-26597 20040419
AB KR 2005101388 A UPAB: 20061027
   NOVELTY - A microorganism capable of degrading dioxins Pseudomonas
   sp. PH-03 is provided to economically and efficiently purify and
   recover environment.
   DETAILED DESCRIPTION - A microorganism capable of degrading
   dioxins and dioxin-like materials Pseudomonas sp. PH-03 (KCTC
   10614BP) is provided, where Pseudomonas sp. PH-03(KCTC 10614BP) is
   isolated from the soil around disposal facilities. A microbial
   composition for degrading dioxins and dioxin-like materials comprises
   Pseudomonas sp. PH-03 (KCTC 10614BP). A method for degrading dioxins
   and dioxin-like materials comprises Pseudomonas sp. PH-03 (KCTC
   10614BP) .
```

Reminder: NCBI/EMBL/DDBJ recently expanded coverage to include KIPO (KR) patents.

GenomeQuest Solution: Software Plus Content

- Search: BLAST, GenePAST, Fragment, Motif
- Content
 - GQ-Pat: automated feeds, continuously updated
 - GeneSeq: requires subscription from Thomson Reuters
 - GenBank, UNIProt, RefSeq, Genomes, Entrez Gene etc..
- Analysis
 - Filter, Group, Sort Results
 - Create Reports
 - Export to BizInt
 - DGENE exclude report

Automated Feeds from Asia-Pacific

- In place
 - Japan: 37,222 documents, 2,519,157 sequences; updated biweekly
 - Korea: 6,412 documents, 168,562 sequences; updated bi-weekly
- Planned for 2009
 - China: Research found 10,106 documents with about 122,000 sequences
- India, Australia and Taiwan to follow

Conclusions

- Asian coverage at NCBI/EMBL/DDDBJ was enhanced to include KIPO during Q1 2008
- DGENE/DWPI and REGISTRY/CAplus cover a wide variety of Asian patent authorities
- USGENE can retrieve additional Asian records via U.S. patent family member relationships
- GenomeQuest GQ-PAT incorporates the Asian coverage provided by NCBI/EMBL/DDDBJ
- A thorough Asian patent sequence search must include all available patent sequence databases

- *Sequence Searching on STN* modular workshop
www.fiz-k.com/bostonsequenceworkshop
 - Sequence Code Match (SCM) searching
 - DGENE, USGENE, PCTGEN content and searching
 - CAS REGISTRY and REGISTRY BLAST
 - Multifile searching using USGENE and DGENE
- USGENE resources, reference materials and FAQ
www.sequencebase.com
- CAS REGISTRY sequence coverage and resources
www.cas.org/support/stngen/stndoc/sequences.html

STN[®]

Coverage of Asian authorities in patent
sequence databases

Robert Austin – FIZ Karlsruhe