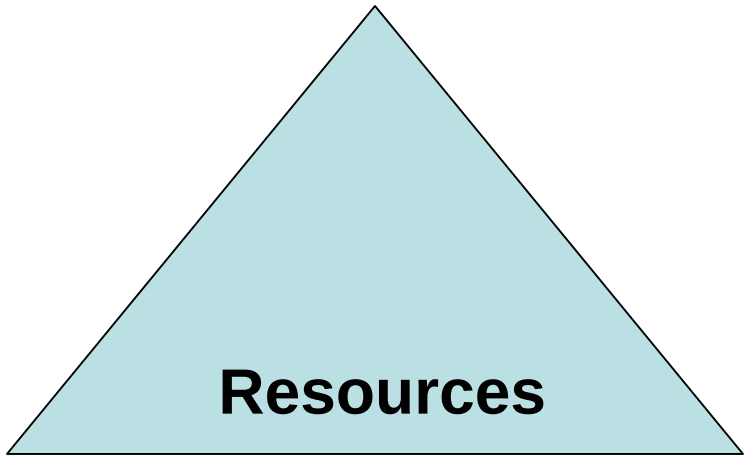
PatentInformatics

Rouse RJ, Castagnetto J, Niedner RH.
PatGen - a consolidated resource for searching genetic patent sequences.
Bioinformatics. 2005 Feb 22;
PMID: 15728120

PatGen DB



Bioinformatic

- NCBI
- EBI
- DDBJ

Patent

- EPO
- WIPO

PatentInformatics

The business of PatentInformatics …

- Is a *HTS < Resources* subsiduary

The business of HTS Resources …

- High Throughput Screening Consulting firm

- Supporting products related to:
  - Microarrays
  - Microfluidics
  - Micromachining
  - Informatics

Patent information is important

- Necessary for technology based businesses to secure venture capital
- Makes the R and D process more efficient
- Patents can be obstacles for market penetration

Patent information is voluminous

Approximately at the USPTO …

- 900 applications are filed a day at the USPTO
- 60 new biotech patents are issued every week
- Patent pdf docs for November 2004 comprised 12 gigs of data

Why was PatGen DB created

- To assist in designing new assays
- To establish a data mining infrastructure
- Supplement for patent consultation
- Help standardize this form of data
- Use the dataset for multiple types of computational analysis (People are patenting BLAST and microarray data!)

PatentInformatics

Public patent bioinformatic resources are inconsistent across different sources

EBI

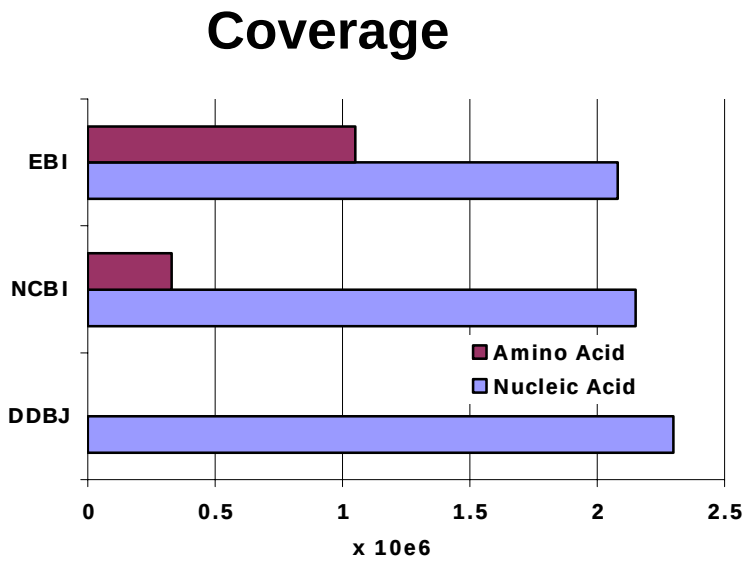
- Nucleic Acids - EMBL
- Amino Acids - EMBL

NCBI

- Nucleic Acids - Genbank
- Amino Acids - FASTA

DDBJ

- Nucleic Acids - Genbank
- Amino Acids - not available



Note: Where do PNAs fall?

Public patent bioinformatic resources lack sufficient patent information

Sequence data contains:

- **Patent title**
- **Inventors names**
- **Applicants**
- **Sequence**
- **Sequence description**

Sequence data lacks:

- **Current family and legal data**
- **Abstracts**
- **Full text**
- **Patent classification codes**

PatentInformatics

PatGen DB is a consolidated resource

PatGen DB

PatGen DB's PubMed Citation

Current Database Holdings - Sequences: 5744678 | Patent Docs: 68569

PatGen is a consolidated database that has been compiled for the purpose of studying legal and bioinformatic issues regarding genetic sequence patent publications. Both issued patents and pending applications can be searched. Through PatGen, current patent legal and family status can be accessed through the European Patent Office Open Patent Service.

PatentInformatics currently maintains the following functionalities:

PatGen DB Lib - This is a basic keyword search tool that enables the searching of bibliographic data. The user can search for the title, abstract, inventor, applicant and date.

PatGen DB Tax - This tool has been set up to enable the searching of patent genetic sequences based on the known sequence taxonomy. This is useful when wanting to know, for example how many patents disclose Fowlpox virus sequences.

PatGen DB Blast - Basic Local Alignment Search Tool (BLAST) is used to perform sequence searching. In this case patents can be searched through the matching of related genetic sequences.

PatentInformatics

Keyword search tool

PatGen DB

Current Database Holdings - Sequences: 5744678 | Patent Docs: 68569

This is a boolean search tool where the user can enter search tokens (i.e., Snake) in one field (i.e., **Title:**) or any number of fields (i.e., **Title:** and **Abstract:** and **Applicant:** and **Inventors:**). The more search tokens you submit the more specific the search is (lower number of results). The dataset corresponding to each search field must contain all search tokens but the tokens do not have to appear in any particular order.

The exception is the **Patent:** field where in this case sequences are retrieved when submitting a patent publication id. This field can not be used when doing multiple field searches.

For your interest BIOS (Biological Innovation for Open Society) hosts a full text biotech patent database. Check that out if you want a more generalized biotech patent search. If you want to search general patent databases, try the European Patent Office's Epacenet or the German Patent Office 's Depatisnet.

Please email us at info@patentinformatics.com if you have any problems with using this resource. Full text searching is available as part of our PatGen DB Server product. We also offer customized searching services.

Title: (i.e., thrombin inhibitors)

Abstract: (coagulation)

Inventors: (i.e., Jane Doe)

Applicant: (i.e., Pfizer)

Patent: (WO9119802)

Date (>=): 01 Jan 1970

Submit

PatentInformatics

Taxonomy search tool (51747 different records)

PatGen DB - Taxonomy

Search based on alphabetical list:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

These are direct links to some of the organisms commonly used in molecular research projects:

Search patents

Arabidopsis thaliana	Hepatitis C virus	Pneumocystis carinii
Bos taurus	Homo sapiens	Rattus norvegicus
Caenorhabditis elegans	Magnaporthe grisea	Saccharomyces cerevisiae
Chlamydomonas reinhardtii	Mus musculus	Schizosaccharomyces pombe
Danio rerio (zebrafish)	Mycoplasma pneumoniae	Takifugu rubripes
Dictyostelium discoideum	Neurospora crassa	Xenopus laevis
Drosophila melanogaster	Oryza sativa	Zea mays
Escherichia coli	Plasmodium falciparum	

Name
Pichia pinus
Pichia sp
Pichia stipitis
Picornaviridae sp
Picramnia pentandra
Picrophilus oshimae
Piedraia hortae
Pieris brassicae
Pieris rapae
Pig endogenous retrovirus
Pig-tailed macaque parvovirus
Pilocarpus heterophyllus
Pimephales promelas

PatentInformatics

BLAST Tool

In order to make a search the query sequence must be in fasta form for example:

```
> AX109268 | WO0123604 | A/1
CAAACTCGTGAGCACATCCTTCTTTCTCGTCAGGTAGGTGTACCTTACATCATCGTATTC
TTAAACAAATGCGACCTTGTTGATGACGAAGAATTACTTGAATTAGTAGAAATGGAAGTA
CGTGAACTTCTTTCTACTTATGACTTCCCAGGTGATGACACTCCAGTAATCCGTGGTTCA
GCTCTTGCAGCGCTTAACGGTGAAGCTGGTCCTTACGGTGAAGAATCAGTTCTTGCTCTT
```

Please contact us at info@patentinformatics.com if you have any questions. A IUPAC reference table is provided.

Select Program:

BlastN

Enter Sequence: (FASTA format only)

Reset

Submit

Note: Because of the limitations on our on-line computational resources, this service might fail or not be available at times, if that happens, try again later.

Result [6]: Seq info - WO0123604 A/147 AX109414

```
Evalue: 1.01229e-41
Identities: 158/182 (86.8%)
Bit score: 174.94

CAAACTCGTGAGCACATCCTTCTTTCTCGTCAGGTAGGTGTACCTTACATCATCGTATTC
|||||||||||||||||||  |   ||||||||| ||||| |||||||||||||||||||
CAAACTCGTGAGCACATCCTGTTAGGTCGTCAGGTTGGTGTTCCTTACATCATCGTATTC

TTAAACAAATGCGACCTTGTTGATGACGAAGAATTACTTGAATTAGTAGAAATGGAAGTA
 | ||||||||  ||| | ||  |||||  |||||  |||||  ||||||||  ||||||||||
CTGAACAAATGTGACATGGTAGATGATGAAGAGTTACTGGAATTAGTMGAAATGGAAGTT

CGTGAACTTCTTTCTACTTATGACTTCCCAGGTGATGACACTCCAGTAATCCGTGGTTCA
||||||||||| |||   || || ||||||||||||||||||||||||||||||||||||
CGTGAACTTCTGTCTCAGTACGATTTCCCAGGTGATGACACTCCAGTAATCCGTGGTTCA

GC
||
GC
```

Free service includes BLASTN and BLASTP, the subscription service includes BLASTX, TBLASTX and TBLASTN.

PatentInformatics

BLAST Considerations many sequence submissions use complicated IUPAC codes

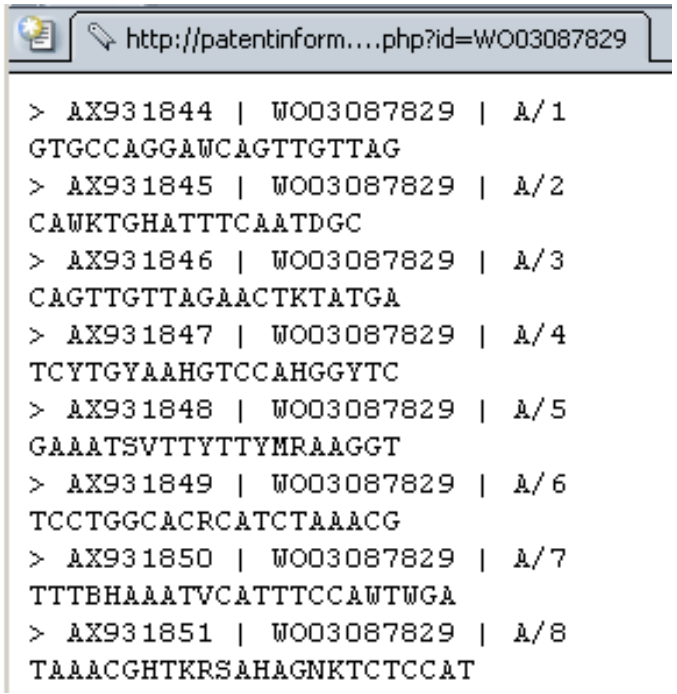
IUB/IUPAC Genetic Sequence Reference

IUB/IUPAC nucleic acid codes

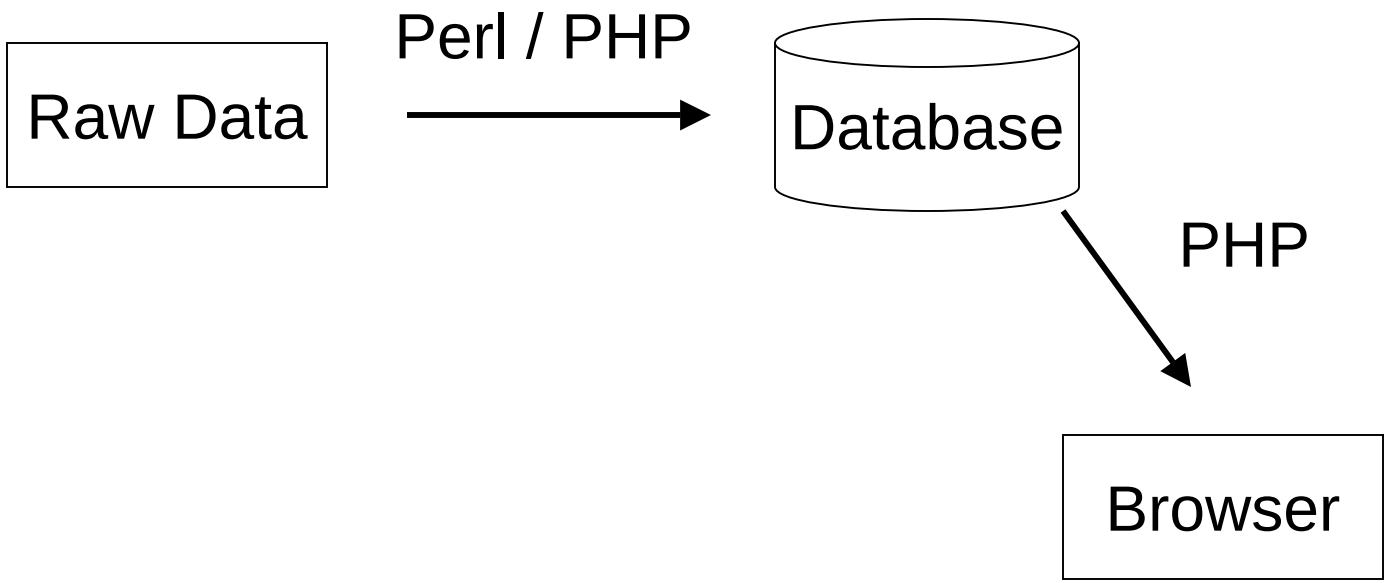
A - Adenine
C - Cytosine
G - Guanine
T - Thymine
U - Uracil
M - A or C (amino)
R - A or G (purine)
W - A or T (weak)
S - C or G (strong)
Y - C or T (pyrimidine)
K - G or T (keto)
V - A or C or G
H - A or C or T
D - A or G or T
B - C or G or T
N - A or G or C or T (any)

Standard IUB/IUPAC amino acid codes

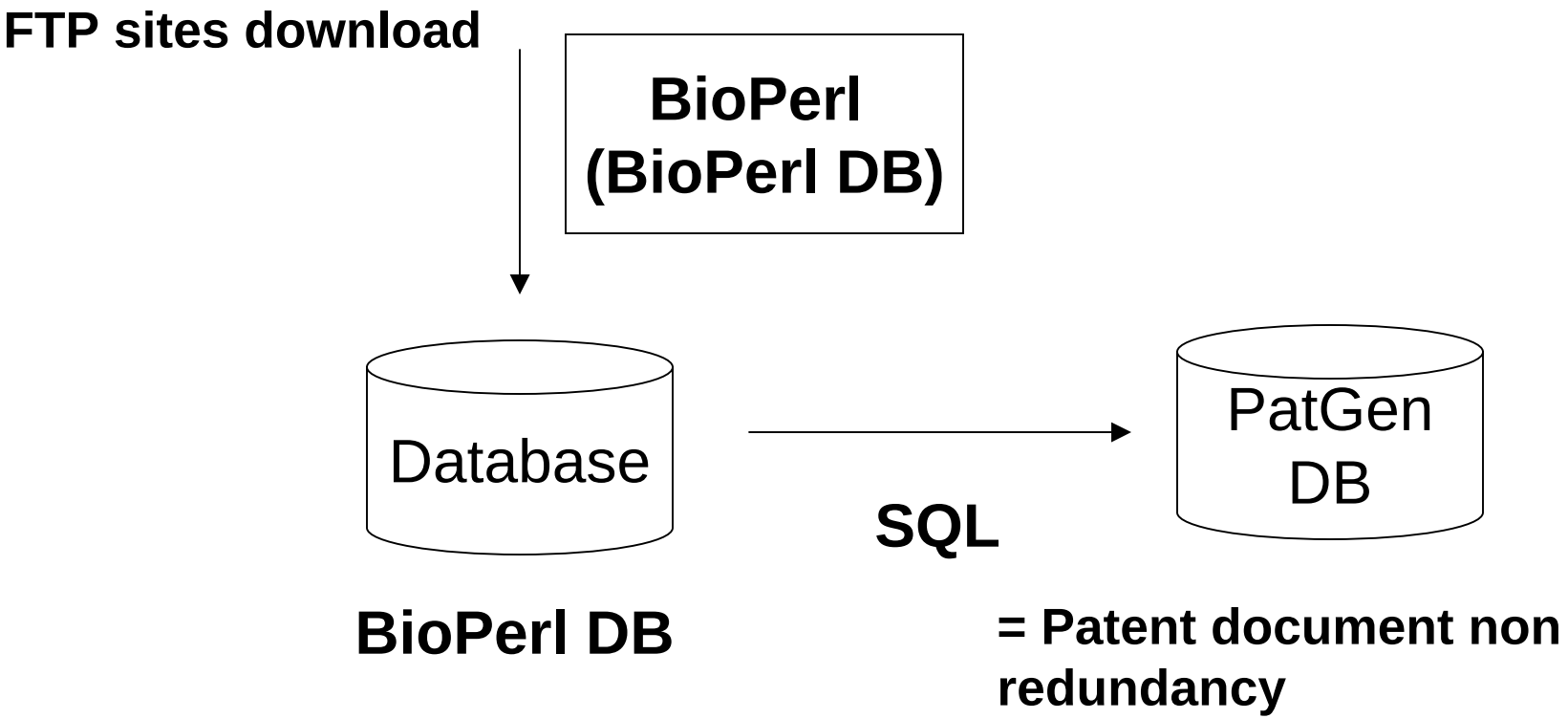
A - Alanine - Ala
B - Aspartic acid or Asparagine - Asx
C - Cysteine - Cys
D - Aspartic acid - Asp



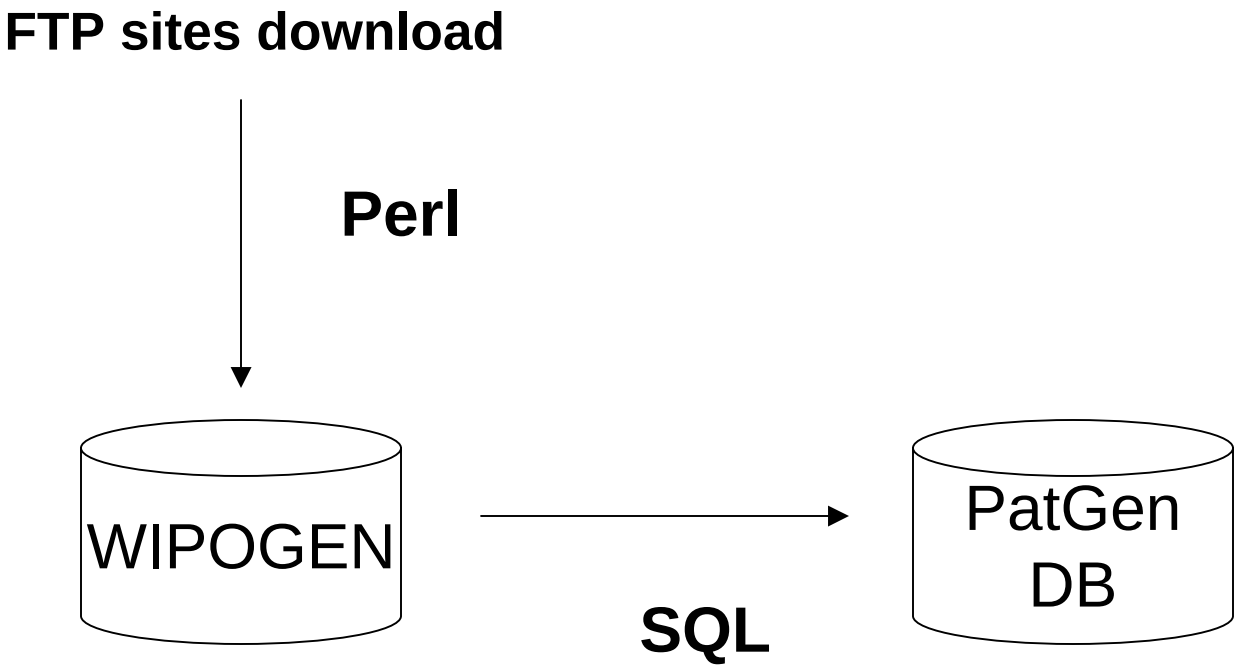
Development of PatGen DB



How we consolidate bioinformatic data from the NCBI, EBI and DDBJ



How we get WIPO sequence data



PatentInformatics

How we get patent data

```
- <OpenPatentServices ACTION="Biblio">
  - <WORLDPATENTDATA>
      <BIBLIO SEED="US5859227" SEED_FORMAT="E" SEED_TYPE="PN"/>
    </WORLDPATENTDATA>
  </OpenPatentServices>
```

European Patent Office
Open Patent Service

```
- <WORLDPATENTDATA>
  - <BIBLIO Seed="US5859227">
    - <SDOBI>
        <B111EP DATE="19990112 ">US5859227</B111EP>
        <B131EP>A</B131EP>
        <B211EP DATE="19970220">US19970803260</B211EP>
        <B311EP DATE="19970220;">US19970803260</B311EP>
        <B311EP DATE="19960731">US19960690010</B311EP>
      - <B510 TYPE="EPC">
          C12Q1/68 (N); C12Q1/68+527/119 (N); C12Q1/68+527/125 (N); G01N33/53F (N); G01N33/68 (N)
        </B510>
        <B510 TYPE="IPC">C07H21/02 ; C07H21/04 </B510>
      - <B542 TYPE="TI">
          RNA sequences which interact with RNA-binding proteins
        </B542>
      - <B570EP>
          Disclosed is a method for identifying possible binding sites for RNA binding proteins in nucleic acid sequences, and confirming the identity of st
          prospective binding sites by detection of interaction between the prospective binding site and RNA binding proteins. Also disclosed are specifi
          sites of RNA binding proteins which have been identified using this method.
        </B570EP>
        <B711EP>BEARSDEN BIO INC (US)</B711EP>
        <B721EP>GIORDANO TONY (US);</B721EP>
        <B721EP>BEACH DEBORAH L (US);</B721EP>
        <B721EP>JACOBS DAVID (US)</B721EP>
      </SDOBI>
    </BIBLIO>
  </WORLDPATENTDATA>
```

PatentInformatics

PatGen DB records

Highly conserved genes and their use to generate probes and primers for detection of microorganisms

WO0123604 - Espacenet

Tabulated sequence: list

Display FASTA (nucleic acids are displayed first): format

Current families, bibliographies and legal status: ops data

Inventors:	Bergeron M.G., Boissinot M., Huletsky A., m Nard C., Ouellette M., Picard F.J., Roy P.H.
Applicants:	Infectio Diagnostic (I.D.I.) INC. (CA).
Date:	05-APR-2001
International Patent Code:	C12Q1/68 ; C07K14/00 ; C12N15/63 ; C12N5/10
European Patent Code:	C12Q1/68M10
Abstract:	Four highly conserved genes, encoding translation elongation factor Tu, translation elongation factor G, the catalytic subunit of proton-translocating ATPase and the RecA recombinase, are used to generate a sequence repertory or bank and species-specific, genus-specific, family-specific, group-specific and universal nucleic acid probes and amplification primers to rapidly detect and identify algal, archaeal, bacterial, fungal and parasitical microorganisms from specimens for diagnosis. The detection of associated antimicrobial agents resistance and toxin genes are also under the scope of the present invention.

PatentInformatics

Sequence browsing in tabulated list format

Displaying rows 1 - 25 of 2283.
< prev next >

Accession	Version	Length	Description	Taxonomy	Seq Type
AX109268	A/1	750	Sequence 1 from Patent WO0123604.	Acinetobacter baumannii	nucleic acid
AX109269	A/2	826	Sequence 2 from Patent WO0123604.	Actinomyces meyeri	nucleic acid
AX109270	A/3	835	Sequence 3 from Patent WO0123604.	Aerococcus viridans	nucleic acid
AX109271	A/4	827	Sequence 4 from Patent WO0123604.	Achromobacter denitrificans	nucleic acid
AX109272	A/5	823	Sequence 5 from Patent WO0123604.	Anaerorhabdus furcosa	nucleic acid
AX109273	A/6	825	Sequence 6 from Patent WO0123604.	Bacillus anthracis	nucleic acid
AX109274	A/7	829	Sequence 7 from Patent WO0123604.	Bacillus cereus	nucleic acid

PatentInformatics

Profile of individual sequence submission

JP2002176989 - **Protein specific to magnetic fine particle membrane**

BD162146

Patent reference:	A/6
Description:	Protein specific to magnetic fine particle membrane. OS Artificial Sequence PN JP 2002176989-A/6 PD 25-JUN-2002 PF 14-DEC-2000 JP 2000381015 PI TADASHI MATSUNAGA,HARUKO TAKEYAMA,YOSHIKO OKAMURA PC C12N15/09,A61K9/127,A61K47/42,C07K16/12,C07K19/00,C12N1/15, PC C12N9/16, PC C12P21/08,G01N33/53,G01N33/543,G01N33/573,G01N33/577//(C12N15/ PC 09 PC ,C12R1:01), PC (C07K16/12,C12R1:01),(C07K19/00,C12R1:01),(C12N1/15,C12R1:01), PC (C12N9/16,C12R1:01),C12N15/00,(C12N15/00,C12R1:01) CC Description of Artificial Sequence:Forward Primer 2 FH Key Location/Qualifiers FT source 1..27 FT /organism="Artificial Sequence"
Taxonomy:	synthetic construct
Length:	27

Sequence:

GAATTCATGGCCGCCAAGCAGACTGAG

Download publication sequences in FASTA format

```
> BD162141 | JP2002176989 | A/1
AAGGSGGTCATCGTGCGGGCCATGCCMAATACCCCGGCGGCGGTGCGGCGCGGCATCACC
GTCTGTGCGCCGGGGAGGCCKTGCCCGTGTCGCCCGTGAGCTTTGCCAGTCGCTGCTGGA
AGCGGTGGGCRGTGGGCTGGGTCGATGACGAGGGCCTGATGGACGTGGTCACSCGTCTCC
GGCTCCGSCCCCGCCTACATCTTCCTCCTGGCCGAGGCCATGGAGGCCGCCGGTCTGGCC
CAGGGGCTGCCCCCGCCCTGGCCGAGCGTCTGGCCCGTGCCACCGTGGSCGGGGCCGGCG
AATTGCTGSGCTGTCCGCCGAACCCGCCGAGCAACTGCGCAAGAACGTCACCTCGCCGGG
CGGCACCACAGCGGCGGCCCTGTCGGTGCTGATGCTGAAAGCCACGGCATTCCCAGCCTG
ATGACCGAAGCGGTGGCTGCTGCCACTCGCCGAGGACGGGAACTTGCGGGCTAGGCGCTT
CGTCAGAGGCGGTACCTATTTCGATATTAGATATTGAGTTGCGTATGACTCCGTTTGACT
CGAAGCCGCCCCCGCGTCTATCTTGCTGCACCGCAACATAAGACCCCCGGGTTGGAGGAA
TAACATGGCCGCCAAGCAGACTGAGCAGTTCTTTGATTTCGACGTCGCCAAGTATCTGGG
CGATTTCAAGGTTCCCGGCGTGGATGTCGAGACCATCGTCGCCAACCAGCGCAAGAACAT
CGAAGCGCTGACCCAGGCGAACAAGCTGGCTTTCGAGGGCCTGCAGAACGTGGTCAAGCG
TCAGGTCGAGATCCTGCGCCAGACCATGGACGAGGTTGCCCAGGTCTCCAAGGATTTCGC
CGAGCCCGGCTCGCCCCAGGGCAAGGCCGCCAAGCAGGCCGAGTTCGCCAAGGATGCCTT
CGAGCGCGCCCTGRCCAACGCCCGTGAGCTGGCCGAGATGATCGCCAAGGCCAATTCCGA
GGCTTTCGACCTGCTGAACAAGCGYTTCACCCAGAGCCTGGACGAGGCCCGCGAGGTCTT
CACCAAGGCCGGCAAGAAGTAAGCCTTCCGTTCATTCGGAACGCTGTCGGCGGCCGCTCC
TGAAAAGGGGCGGCTNTTTGATTCCGGCCTGAATTGGGCGCTCTACCGTCCCTCGATCAG
GGCCAGAAGCGGCGCCCATTCGGCGGCGTAGGACTTGGAACGGCGGTCATTGGCCTCGAT
CACCGTCGAGCCGGCGGCGGCCACCGCGGTATAGA
> BD162142 | JP2002176989 | A/2
CATAAGCAGACCGAGCAGTTCTTCGA
> BD162143 | JP2002176989 | A/3
TTGGCCTGGGTCAGGGCCTCGATGTT
> BD162144 | JP2002176989 | A/4
CGCTGGTTGGCGACGATGGTCTCGACATCC
```

PatentInformatics

Access current family and legal data through OPS

Doc. Number	Date	Appl. No.	EPOQUE No.	Prio. No.	Prio. Date
JP 2002176989A	20020625	JP 2000381015A	JP2002176989	JP 2000381015A	20001214

Classifications:

- EPC: C07K14/195; C12N9/14
- IPC: C12N15/09; A61K9/127; A61K47/42; C07K16/12; C07K19/00; C12N1/15; C12N9/16; C12P21/08; G01N33/53; G01N3 G01N33/573; G01N33/577; C12N15/09; C12R1/01; C07K16/12; C12R1/01; C07K19/00; C12R1/01; C12N1/15; C12R1/01; C12R1/01

Applicant: JAPAN SCIENCE & TECH CORP

Inventors:

- MATSUNAGA TADASHI;
- TAKEYAMA HARUKO;
- OKAMURA YOSHIKO

Titles:

- TI: MAGNETIC PARTICULATE MEMBRANE-SPECIFIC PROTEIN

Abstract

- PROBLEM TO BE SOLVED: To provide a magnetic particulate membrane-specific protein Mms16 derived from a magnetot bacterium (Magnetospirillum sp.) strain AMB-1, a DNA encoding the same, and sandwich immunoassay methods and pharma

Doc. Number	Date	Appl. No.	EPOQUE No.	Prio. No.	Prio. Date
JP 2002176989A	20020625	JP 2000381015A	JP2002176989	JP 2000381015A	20001214

Doc. Number	Date	Appl. No.	EPOQUE No.	Prio. No.	Prio. Date
US 2004048289A1	20040311	US 45034603A	US2004048289	JP 0108918W	20011011
				JP 2000381015A	20001214

Doc. Number	Date	Appl. No.	EPOQUE No.	Prio. No.	Prio. Date
WO 0248359A1	20020620	JP 0108918W	WO0248359	JP 2000381015A	20001214

Legal Events:

2002-06-20 - AK +DESIGNATED STATES Kind Code of Ref Document A1

2003-01-03 - DFPE REQUEST FOR PRELIMINARY EXAMINATION FILED PRIOR TO EXPIRATION OF 19TH MONTH FROM PRIO

Families and Legal example

Features of PatGen DB commercial service

- Custom BLAST analysis
- Full text data access
- Custom database interfaces
- Sessions for user

Future Directions

- **Help change paradigm regarding the business of patent information (Hey its not the data but the service!)**

- **PatIndex commercial service – a generic patent database platform**

- **Implement patent information issues in experimental design through *HTS < Resources***