Patent Search For Engineers And Lawyers

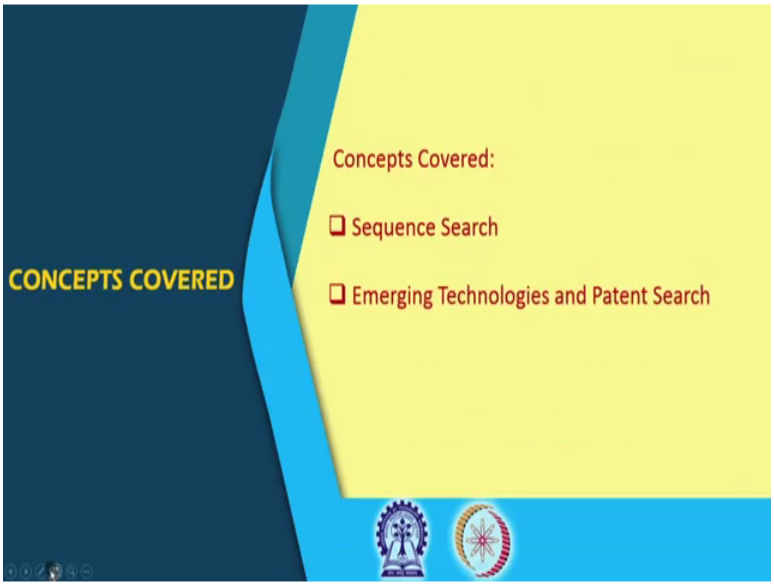
Prof. M. Padmavati
Rajiv School Of Intellectual Property Law
Indian Institute of Technology, Kharagpur

Lecture – 14

Hands on Patent Search – Sequence Search

Welcome to the lecture on the Hands on Patent Search. In the earlier lectures we have looked at the aspects of search through the different databases and also the structure search.

(Refer Slide Time: 00:36)



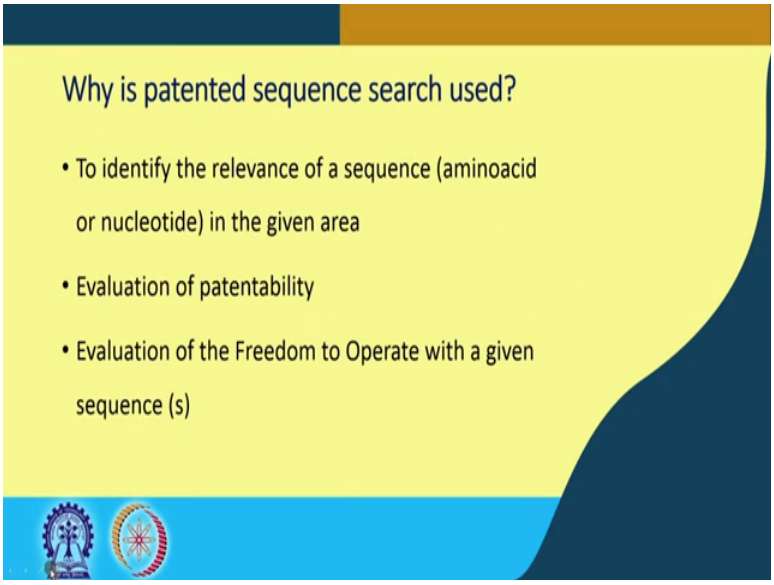
CONCEPTS COVERED

Concepts Covered:

- ❑ Sequence Search
- ❑ Emerging Technologies and Patent Search

In this particular lecture, we will be looking at the aspects of sequence search and also search conducted in relation to emerging technologies.

(Refer Slide Time: 00:46)



Why is patented sequence search used?

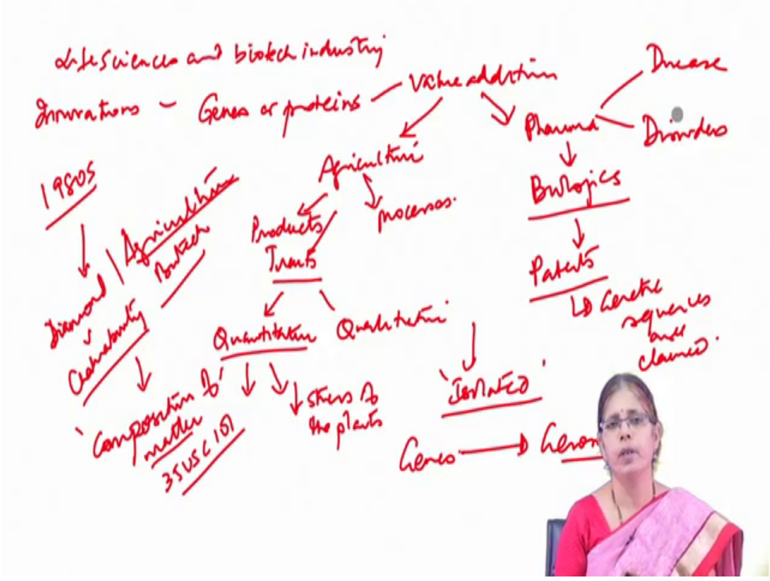
- To identify the relevance of a sequence (aminoacid or nucleotide) in the given area
- Evaluation of patentability
- Evaluation of the Freedom to Operate with a given sequence (s)

One of the important aspects of search is the patented sequence search. This sequence search is very important in industries like the life sciences and agriculture industry where the entire basis of innovations coming on are in relation to protein sequences and nucleotide sequences.

So, patented sequence search is done for various reasons. One it is done to understand the relevance of a particular sequence in relation to an area. The second purpose with which patented sequence search is done is fundamentally to assess the novelty of a sequence. So, it is imperative for any new sequence that is identified to know whether there is a novelty with respect to that sequence. So, it is very important to look at the patentability aspect. Patented sequence of search are also done in order to understand the freedom to operate for a given set of sequences.

So, let us understand a little bit of basics in this particular area in relation to what are gene sequences and what are protein sequences and how are they relevant for the certain industries which operate in this particular area.

(Refer Slide Time: 02:03)



Let us understand certain basic aspects of how the life science and the biotech industry work. Now innovations can come in various ways in this industry where they are working on certain products and processes and one predominant area which the agricultural industry and the biotechnology industry which includes in certain instances they biopharma industries, the work on identifying new genes and proteins which are relevant for some value addition. So, we are looking at genes or proteins for value addition.

Now, in the area of agriculture, they are to for improved products. Here we talk about traits and in the many other instances it is about improving processes. For instance, we are talking about let us say an improving nitrogen uptake there could be a set of genes which are involved in improving the nitrogen uptake. Now when we are looking at traits, traits can be of two types; quantitative and then qualitative.

So, the qualitative traits are those traits which are inherent in terms of the providing the qualitative perspective in relation to a particular plant. So, here we are talking about a particular stay green is one qualitative trait in which the plants continue to stay green.

Now in quantitative traits often it is about improving yield in improving the total capacity of the plant. So, they had there are certain genes which are involved in improving yield. So, yield component can come through various ways in terms of just improving the yield in terms of the specific pathways or improving yield indirectly by actually lowering this stress of the plant.

So, the area of agriculture biotechnology has grown into really large dimension from the point of view of today; we have not only gene, but gene sets being taken as the inventions. So, the area of agriculture biotechnology has grown to such tremendous dimensions that it is actually unparalleled. And how has this growth happened it is important for especially those who are from the area of biotechnology and agriculture is that we need to go back to the times of 1980s. This is the time when the very first wave of patents in relation to gean started with the celebrated case of Diamond versus Chakrabarty.

Diamond versus Chakrabarty is one of the its the minor cases in relation to opening up the entire area for biotechnology patenting. In a very simple sense, the case deals with the aspect of whether artificially created microorganisms can they be a subject matter for patent.

So, the invention basically created by Anand Mohan Chakrabarty; Doctor Anand Mohan Chakrabarty was a bacterium which was artificially made with the introduction of four different plasmids. That particular pseudomonas species did not have earlier any of those four different plasmids. So, it was constructed, it is a constructed bacterium and such a constructed bacterium was very efficiently able to dissolve the fatty acids into. So, this was purported to be a very good solution for degradation of oil spills.

Now, this was the patent application filed at the USPTO and initially was rejected by the USPTO because at that point of time living material was not interpreted under the purview of 35 USC 101 35 USC is the patent code of the united states and 35 USC 101 relates to the subject matter that is eligible for patentee. Not earlier did the patent office ever realized that living organisms artificially created could be subject matter for patents.

So, this case brought in a very important change to the entire area of patenting; on an appeal against the decision of the controller, this moved to the court this case moved to the court and the court decided in favor of Doctor Chakrabarty and so, the first patent to be opened up for the area of artificially created organisms to be read under the purview of subject matter is became a very important development.

So, today the area of biotechnology particularly, if you look at the those with gene interventions is read under the expanded context of composition of matter; composition of matter under the purview of 35 USC 101. So, beginning the 1980s and then the area developed a lot of isolated genetic sequences became subject matter for patents and their value and then again the area of pharmaceutical industry. The pharmaceutical industry also realized the value of bio based interventions for creation of value in terms of drugs.

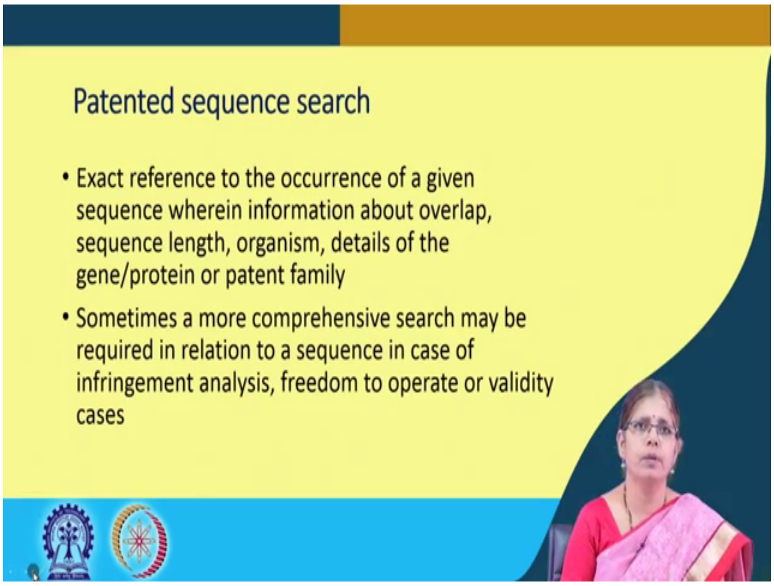
So, the area of biopharma also picked up and today we have a lot of sequences which are bio based which are also subject matter for patents. Today we have recombinant drugs which are an important area for patenting recombinant which are different types of recombinant drugs which also include recombinant vaccines. So, that is the other kind of value addition that you see in relation to the pharma industry where today we talk about biologics which are bio based drugs and this is again opened up a whole lot of patents where a lot of genetic sequences are claimed.

One important consideration that we need to keep in mind is how to make out whether a sequence is non nature. So, essentially the sequence is isolated and its complimentary DNA component is actually the one that is a consideration for patenting. And the various forms of it the various interactions of that particular sequence with others sequences becomes an open subject matter for patenting. So, in all claims in patents in relation to gene sequences, the word isolated the markets; the context of what is nature versus non nature. This fundamental understanding is important for us from the point of view of understanding patenting in relation to genetic sequences.

The very early sequence patents were essentially protein sequence patterns and then of course, you have the with the improvement of DNA sequence techniques and advanced molecular biology techniques. Today you have a whole or range of sequences not only from gene sets, but it is also about genomes such as the importance of this area that today there is also a patent called minimal genome which is the artificially created genome in relation to us one of the smallest microorganism.

So, today we are not just talking about genes and their modifications. We are also looking at genomes and the value that can be created out of the use of genomes, thus the sequence of the genomes is very huge. Today this has brought in definite interest in the area of improving value with respect to disease and disorders. This is again an important area of search which needs a capability development in relation to looking at how sequences are searched.

(Refer Slide Time: 11:44)



Patented sequence search

- Exact reference to the occurrence of a given sequence wherein information about overlap, sequence length, organism, details of the gene/protein or patent family
- Sometimes a more comprehensive search may be required in relation to a sequence in case of infringement analysis, freedom to operate or validity cases

So, the moving on patented sequence search. So, what is it that we can expect when we are looking at a patented sequence search? So, we are looking at possibly the first instance would be a very general search which would mean you are looking at the aspect of the occurrence of a given sequence or a set of sequences in relation to a particular organism.

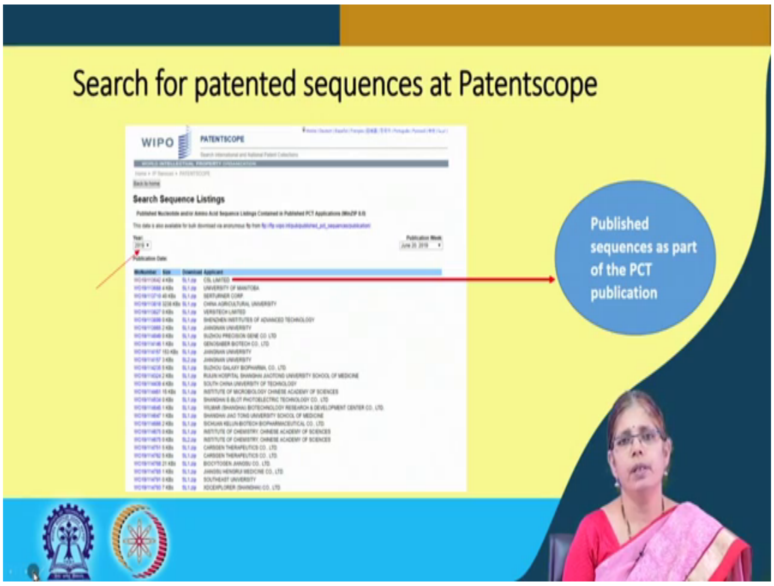
So, what is it that we were looking at here? We are looking at is there a overlap, but with my query sequence with other sequences there are present in the database. In terms of understanding whether other organisms you know already have this kind of sequence which are isolated what is the length of the sequence and does the sequence belong to a set of families of proteins or let us say gene sequences.

So, sequence search can be done for general aspects sequence search sometimes can be very comprehensive from the point of understanding the value of a particular query in relation to a particular patent search option. For instance, I can input a particular query sequence to understand whether there is an issue with respect to an infringement. So, in an infringement analysis you are maybe looking at certain sequences and if there is an overlap with an existing sequence which is already patented.

The other aspect of doing sequence search is to look at when you are improved when you are bringing in the aspect of freedom to operate many a time the gene sequences that you are taking may be novel, but often when gene sequences are implemented in an organism you may need to use certain promoters. It is like a fuel for a engine in very simple sense. So, if a gene has to work, it needs its promoter.

Now the question that one needs to sometimes address is that do we have the freedom to operate with that particular promoter. Now promoters are of different types. So, the question that comes to ones mind is does the particular concern whether is it is an institution or a company have the freedom to operate with that set of promoters. So, therefore, freedom to operate search also is an important aspect which is taken into consideration and in many cases for validity search also one looks at patented sequence information.

(Refer Slide Time: 14:19)

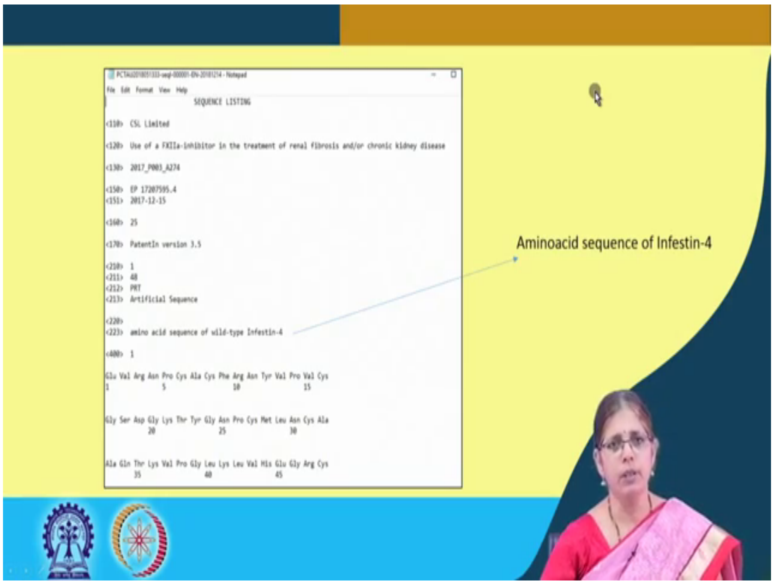


Search for patented sequences at Patentscope

Published sequences as part of the PCT publication

So, what are the various sources available in the public databases in relation to doing a patented sequence search? Today the WIPO patent scope provides and search option for sequences in relation to several different organisms.

(Refer Slide Time: 14:48)



```
                    SEQUENCE LISTING

<110>  CSL Limited

<120>  Use of a FXIIa-inhibitor in the treatment of renal fibrosis and/or chronic kidney disease

<130>  2017_P003_A274

<150>  EP 17207595.4
<151>  2017-12-15

<160>  25

<170>  PatentIn version 3.5

<210>  1
<211>  48
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  amino acid sequence of wild-type Infestin-4

<400>  1

Glu Val Arg Asn Pro Cys Ala Cys Phe Arg Asn Tyr Val Pro Val Cys
1               5                   10                  15


Gly Ser Asp Gly Lys Thr Tyr Gly Asn Pro Cys Met Leu Asn Cys Ala
            20                  25                  30


Ala Gln Thr Lys Val Pro Gly Leu Lys Leu Val His Glu Gly Arg Cys
        35                  40                  45
```

So, if one book looks at the WIPO patent scope, one can actually look at the information that is provided under the WIPO page. So, this is a database which is regularly updated and it has information in relation to sequence search beginning from 2019 to backwards. So, you can search by the year. So, what is the information that you get? The information that is provided is you have the patent document which wants the base document for the set of sequences which are claimed within that particular patent. You also have the details in relation to the assignee in terms of who is the applicant in relation to that particular patenter. So, you have details of the applicant provided, then you also have details in relation to the sequence data itself.

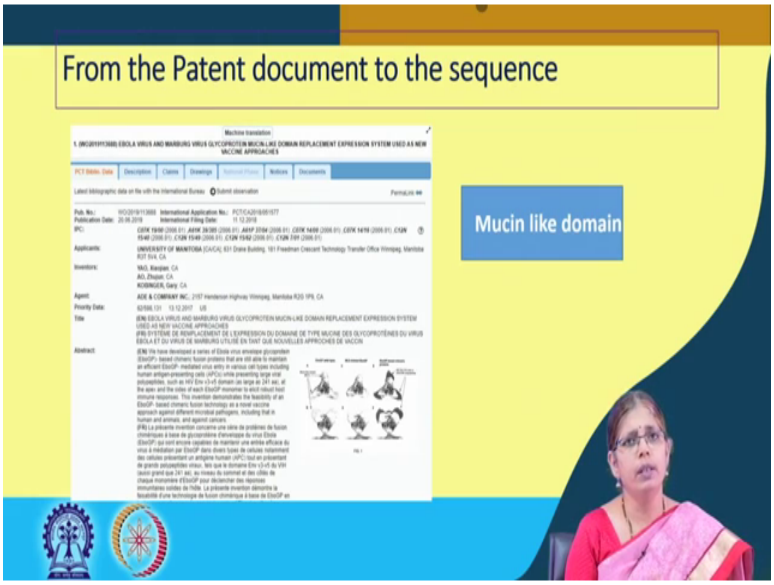
So, the sequence data itself is provided in the downloadable option at each of these links. So, you can have various sizes of sequences available and so, the document size gives you an idea about the amount of sequence information present in that particular data. So, today at the WIPO website using patent scope one can actually go for a sequence listing search the search encompasses the data in relation to the published PCT applications which are available at the in the WIPO database.

So, for instance if we search by the very first sequence that is represented in this in the window. Here is the sequence which is listed which is the use of FX III a inhibitor in the treatment of renal fibrosis and or chronic kidney disease.

So, the application relates to an EP European patent application and the pre RT as you see down in the link is the protein sequence, you can also see the listing it mentions very clearly the artificial sequence. So, what could one expect from a file of this nature? One is essentially the entire aminoacid sequence listing of that particular protein. In this case, it says amino acid sequence of wild type Infestin-4.

Now, you can have different versions of this particular amino acid sequence which could be possibly part of this particular application.

(Refer Slide Time: 17:30)

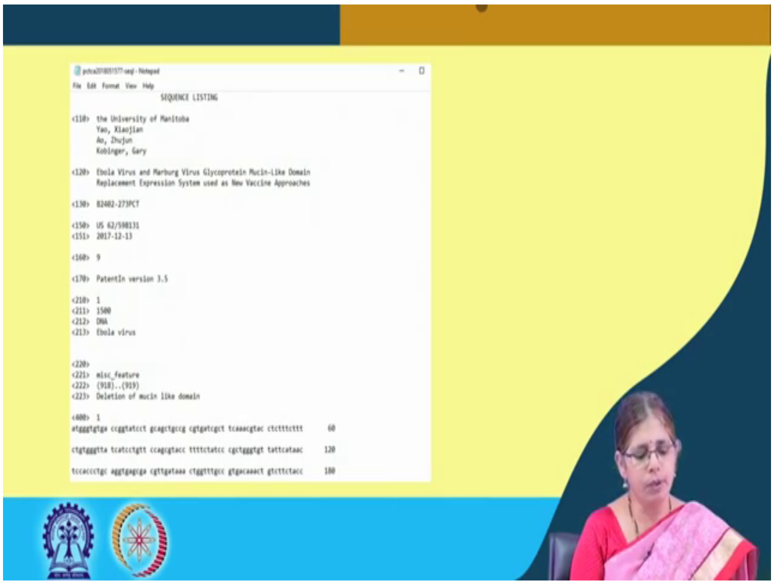


From the Patent document to the sequence

Mucin like domain

We can get a lot of data not only from the sequence information, but we can also go back to the particular patent application and map the value of that particular sequence. So, here is another example of Ebola virus and the subject matter of this patent is about Mucin like domain replacement expression system and its use in developing vaccines. This the applicant for this is University of Manitoba and there are a series of the proteins that are actually being discussed as a part of this particular application. And so, one can actually understand what are the type of sequences claimed in this particular patent application and particularly the focus here is on Mucin like domains.

(Refer Slide Time: 18:32)



```
                    SEQUENCE LISTING

<110>  the University of Manitoba
       Yao, Xiaojian
       Ao, Zhujun
       Kobinger, Gary

<120>  Ebola Virus and Marburg Virus Glycoprotein Mucin-Like Domain
       Replacement Expression System used as New Vaccine Approaches

<130>  82402-273PCT

<150>  US 62/598131
<151>  2017-12-13

<160>  9

<170>  PatentIn version 3.5

<210>  1
<211>  1500
<212>  DNA
<213>  Ebola virus



<220>
<221>  misc_feature
<222>  (918)..(919)
<223>  Deletion of mucin like domain

<400>  1
atgggtgtga ccggtatcct gcagctgccg cgtgatcgct tcaaacgtac ctctttcttt      60

ctgtgggtta tcatcctgtt ccagcgtacc ttttctatcc cgctgggtgt tattcataac     120

tccaccctgc aggtgagcga cgttgataaa ctggtttgcc gtgacaaact gtcttctacc     180
```

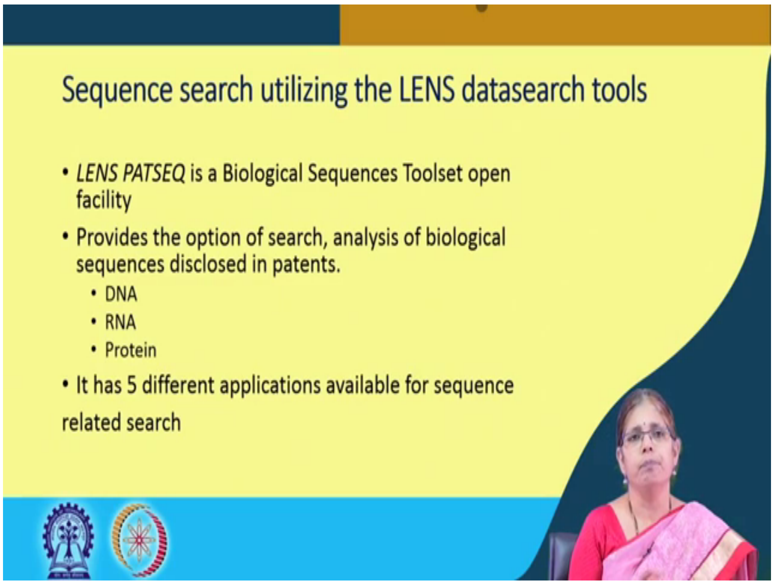
So, here you see in relation to the this particular patent, there is a sequence deposit available as part of the sequence listing. In this case, the DNA sequence is presented. Particularly since the focus is on Mucin like domain and the its approach I am using this to create the novel vaccine constructs. So, here the focus has been on the deletions that are possible in the Mucin like domain to simply put, there are specific domains in DNA sequences which are have value in relation to elevation with respect to disease or disorder. So, sometimes if you delete certain part of the stretch, it can be of improved value.

So, here they purported interest in this particular patent or this in this invention is about deletion of Mucin like domain. So, there are some deletions within this. So, this is the representation of that. One thing that we need to keep in mind when it comes to the biotechnology patents and the those in relation to biopharma is often, if you look at patents you would find that a range of sequences are actually claimed which means for instance let us say, I am claiming a nitrogen reductase; simply speaking nitrogen reductase as one of the genes.

So, there could be different forms of that nitrogen reductase which could be also part of the patent. So, you could have for instance if nitrogen reductase is called NR. You could have NR like one NR like two and a whole battery of the genes claimed in that particular application. So, often if you look at the biotechnology applications, you can find a lot of broad claims.

So, for sequence searches especially what is important is to look at not only the basic sequence of NR, but also the like forms. So, novelty assessment has to be done in relation to each of those. So, going back to the patent application is always useful to identify exactly what are the entire gene sets or the protein sets claimed in that particular application. You can make a listing of that into a downloadable format into your either an excel sheet or a word document and from there you can actually move about individually looking for the novelty of these particular sequences.

(Refer Slide Time: 21:09)



Sequence search utilizing the LENS datasearch tools

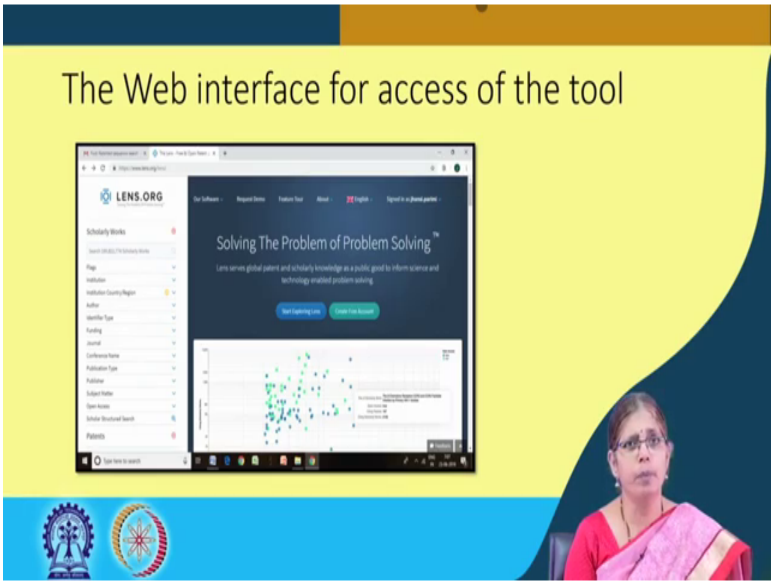
- *LENS PATSEQ* is a Biological Sequences Toolset open facility
- Provides the option of search, analysis of biological sequences disclosed in patents.
    - DNA
    - RNA
    - Protein
- It has 5 different applications available for sequence related search

Another important database in relation to conducting sequence search is the search available at LENS dot ORG. Now the LENS database search tools are provided on an open access format. So, therefore, one can actually given a user id password and register at the database and I get the access for the search option that is available at the LENS data search.

Now the LENS data search provides very enhanced option for doing sequence search and analysis in various ways. The primary aspect of the lens data search is what we call the LENS PATSEQ. Now the LENS PATSEQ is basically a sequence search option available at the lens database the options that are there with respect to what molecules can be searched are three; DNA RNA and protein.

Besides the options of basic search, it also provides five different types of facilities in relation to sequence search that is you can have a extended search done; analysis done. So, let us explore what are the other different aspects of this particular search facility available at the lens database.

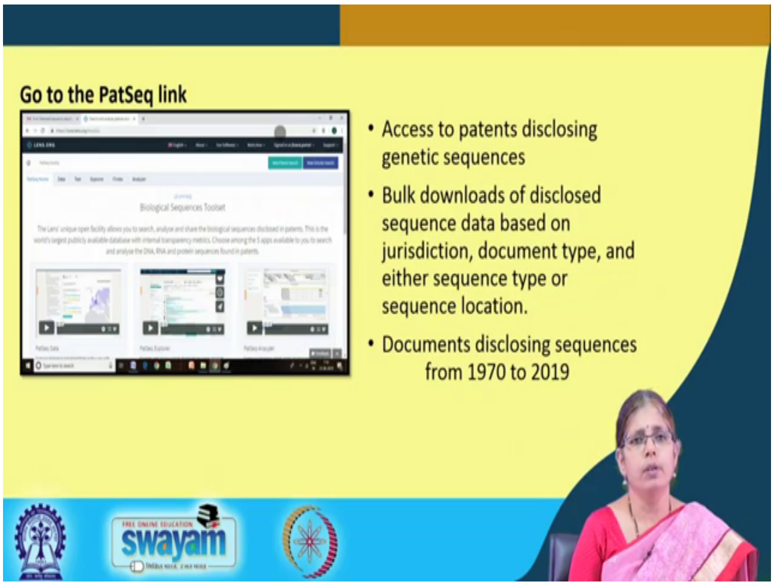
(Refer Slide Time: 22:41)



The Web interface for access of the tool

This is the basic interface for the access of the tool. So, if you go to the option of LENS dot ORG. One can now look at the interface and it provides for actually a very user friendly approach in relation to the search.

(Refer Slide Time: 23:08)

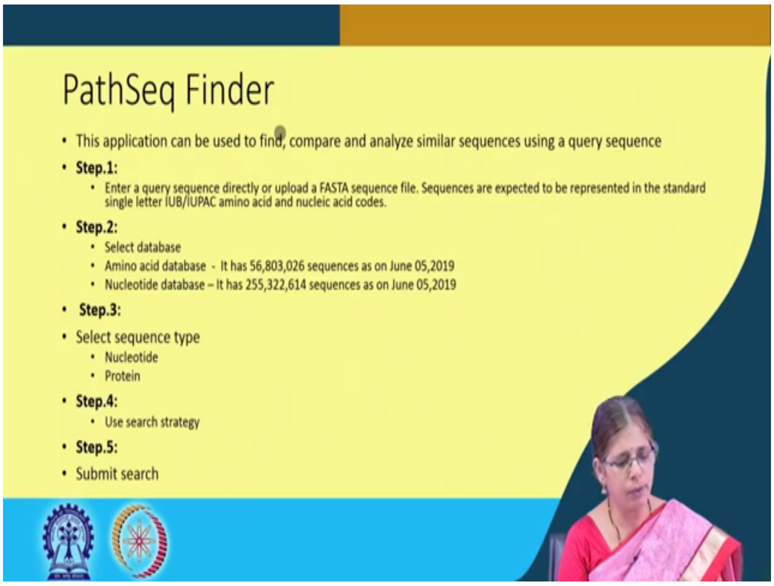


Go to the PatSeq link

- Access to patents disclosing genetic sequences
- Bulk downloads of disclosed sequence data based on jurisdiction, document type, and either sequence type or sequence location.
- Documents disclosing sequences from 1970 to 2019

One of the important links that is available is the PATSEQ link which is the biological sequences toolkit available in the PATSEQ database. And the how does it help? It actually helps to search analyze and share the sequence which are disclosed in the patent applications. So, the option of bulk loading of the data in terms of the looking at the entire data in terms of downloading the entire set of sequences is available which is a very useful feature it has the sequence listing beginning from 1970s to the 2019 as recent. And various documents can be accessed based on the genetic sequence information that is available in the PATSEQ.

(Refer Slide Time: 24:04)



PathSeq Finder

- This application can be used to find, compare and analyze similar sequences using a query sequence
- **Step.1:**
  - Enter a query sequence directly or upload a FASTA sequence file. Sequences are expected to be represented in the standard single letter IUB/IUPAC amino acid and nucleic acid codes.
- **Step.2:**
  - Select database
  - Amino acid database - It has 56,803,026 sequences as on June 05,2019
  - Nucleotide database – It has 255,322,614 sequences as on June 05,2019
- **Step.3:**
- Select sequence type
  - Nucleotide
  - Protein
- **Step.4:**
  - Use search strategy
- **Step.5:**
- Submit search

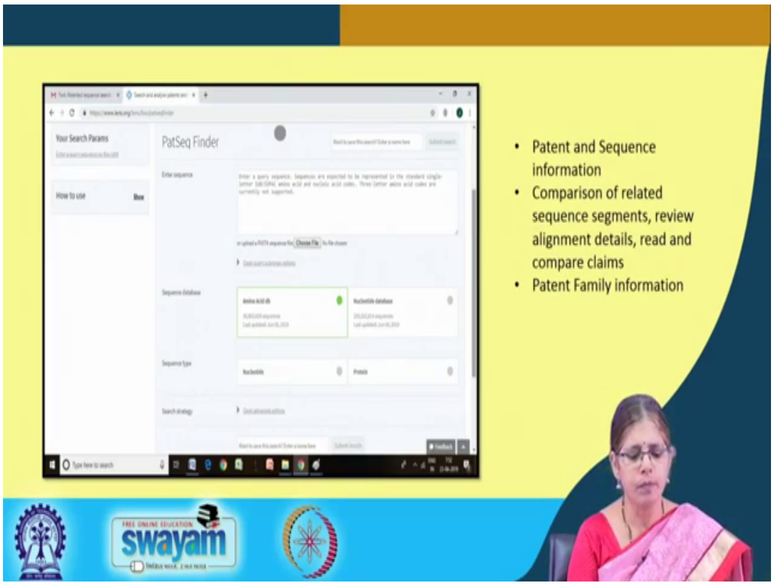
So, let us look at the further features that are available at the PathSeq finder. This particular application is used to actually find sequence. So, one is to look at the finding the sequence for the query sequence that is available, then you can compare the sequence. You can look at the analysis in relation to the sequence data in terms of which are the domains which are actually overlapping.

So, that sort of analysis one can do with PathSeq finder. So, there is a basic steps involved in doing the analysis in relation to PathSeq finder. The first option that is available is basically the query option in which either you directly put the sequence or you can upload a faster file and sequences need to be represented in specific type of formats which is specified in the database and one can actually select the specific type of database.

Now, here you can either select the amino acid database or you can select the nucleotide database. It is also important to take keep in mind the updation of information available at each of these databases and the total list of sequences in the database. The LENS dot ORG basically sources its data from different public databases which are available in the world. So, at the individual patent officers when one is applying for a patent in relation to a biotechnology application or a particular; let us say a bio pharma application, one would actually also upload the sequence file.

So, those entire list of sequences from the individual patent collections of the databases are utilized to populate that into the LENS dot ORG. Then you can look at the third option of where you can select the sequence which is either a nucleotide or a protein sequence and then run the strategy of search.

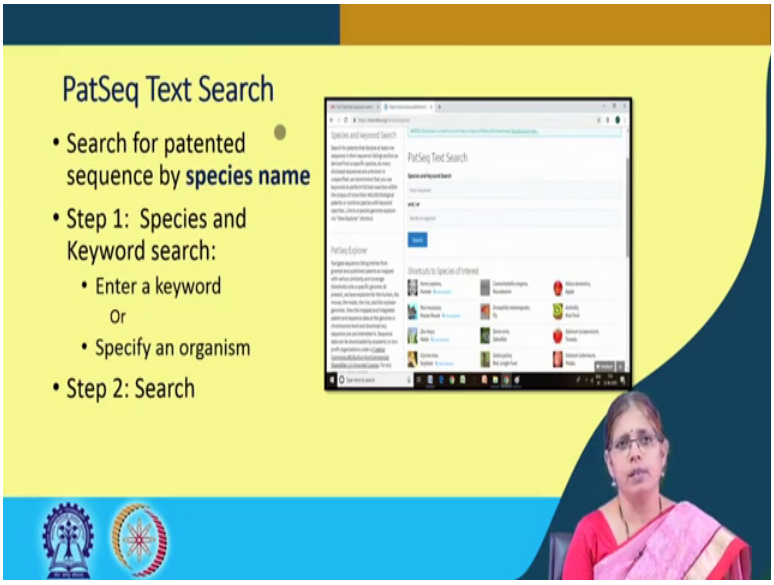
(Refer Slide Time: 26:08)



- Patent and Sequence information
- Comparison of related sequence segments, review alignment details, read and compare claims
- Patent Family information

This is the interface that comes up. So, enter the sequence and then you can actually go for a specific search in relation to certain alignment details in that particular sequence; one can actually also source for patent family information.

(Refer Slide Time: 26:32)



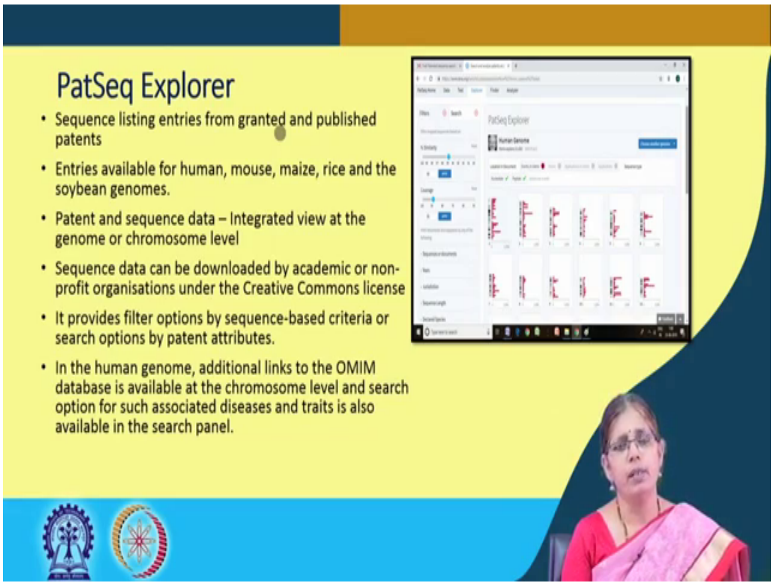
PatSeq Text Search

- Search for patented sequence by **species name**
- Step 1: Species and Keyword search:
  - Enter a keyword

    Or
  - Specify an organism
- Step 2: Search

So, PatSeq is another type of a search or an application available under the lens dot org which is basically for searching of the sequence by species name. So, one can input that into the keyword search window or you can actually search by the name of the organism and then look at the entire set of sequences within that.

(Refer Slide Time: 26:55)



PatSeq Explorer

- Sequence listing entries from granted and published patents
- Entries available for human, mouse, maize, rice and the soybean genomes.
- Patent and sequence data – Integrated view at the genome or chromosome level
- Sequence data can be downloaded by academic or non-profit organisations under the Creative Commons license
- It provides filter options by sequence-based criteria or search options by patent attributes.
- In the human genome, additional links to the OMIM database is available at the chromosome level and search option for such associated diseases and traits is also available in the search panel.

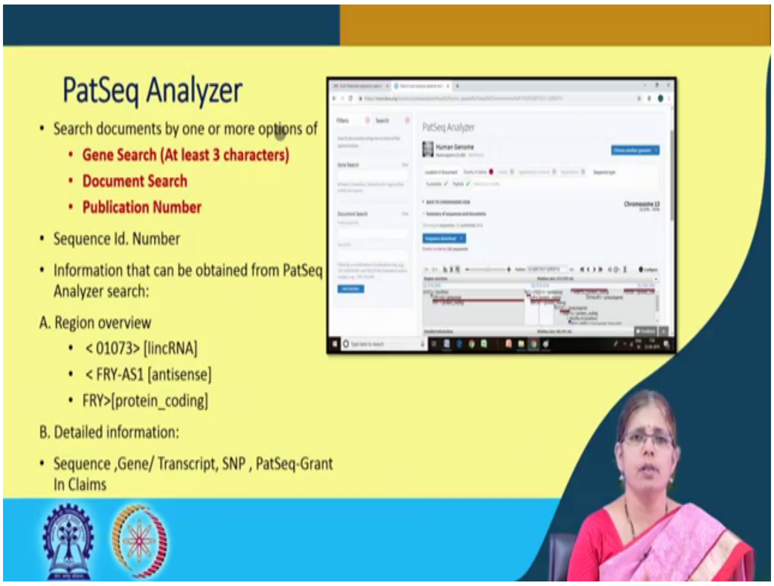
Now, PatSeq has another option of what we call PatSeq Explorer. There entries from different published applications and granted application are granted patents are available.

The sequence data is specifically available for the maize human genome riser the mouse genome and the soybean genome.

The value of this PatSeq explorer is that it provides you a integrated view of the presence of these particular genes at the genome on a chromosome level and. So, that is one good option that is available. Such data can also be accessed using a creative commons license. There are various filter options available for the sequence based criteria. So, one can see a lot of enhanced attributes in relation to this particular database.

Now, in the case of the human genome you have additional links to the OMIM which is where you have the different aspects of access to information available in relation to a particular disease and all of that.

(Refer Slide Time: 28:04)



PatSeq Analyzer

- Search documents by one or more options of
  - **Gene Search (At least 3 characters)**
  - **Document Search**
  - **Publication Number**
- Sequence Id. Number
- Information that can be obtained from PatSeq Analyzer search:

A. Region overview

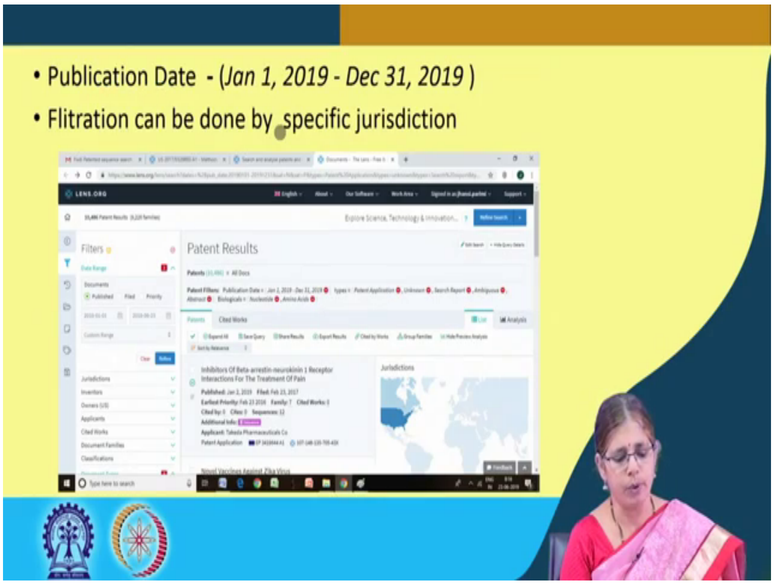
- < 01073> [lincRNA]
- < FRY-AS1 [antisense]
- FRY>[protein_coding]

B. Detailed information:

- Sequence ,Gene/ Transcript, SNP , PatSeq-Grant In Claims

Another application that is available at the LENS dot ORG is the PatSeq analyzer. Now the PatSeq analyzer is important from the point of view of looking at the information in relation to sequences and specifically looking at the aspects of extended aspect of the sequence which are you know possible for analysis. So, you can actually pull out the sequence and drop in the sequence for into another window for further analysis. So, this analysis option is again very useful particularly using the PatSeq analyzer.

(Refer Slide Time: 28:37)

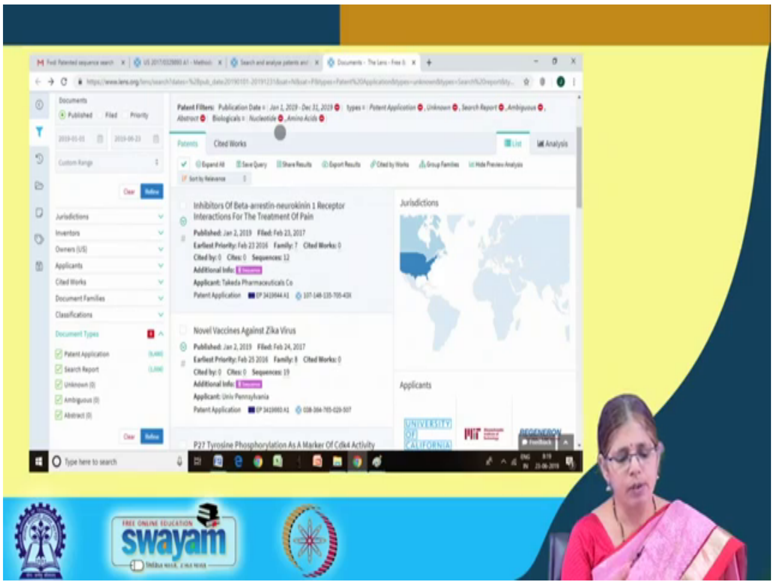


- Publication Date - (*Jan 1, 2019 - Dec 31, 2019*)
- Flitration can be done by specific jurisdiction

And one such example of the how one can use PatSeq is available in terms of the jurisdiction position of the windows where you can so, source of the sequence and also do the analysis in the window just beside the main window. One can actually also look at routinely what are the new gene sequences coming out and that information can be answer accessed from PatSeq.

Particularly, for instance if you are looking at the data between January 2019 to 31 st of December 2019, you will have a continuous listing by numbers of each of these particular sequences listed. Of course, one can also do the data filtration by a specific jurisdiction by a specific country. This is one illustration of the patent results that all documents are sourced for a given set of publication dates.

(Refer Slide Time: 29:38)



Patent Filters: Publication Date = Jan 1, 2019 - Dec 31, 2019, types = Patent Application, Unknown, Search Report, Ambiguous, Abstract, Biologicals = Nucleotide, Amino Acids

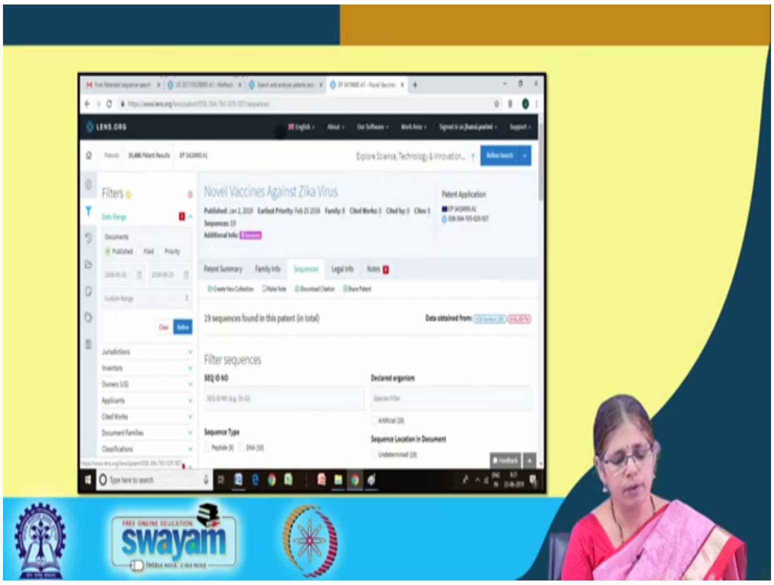
Inhibitors Of Beta-arrestin-neurokinin 1 Receptor Interactions For The Treatment Of Pain
Published: Jan 2, 2019 Filed: Feb 23, 2017
Earliest Priority: Feb 23 2016 Family: 7 Cited Works: 0
Cited by: 0 Cites: 0 Sequences: 12
Applicant: Takeda Pharmaceuticals Co

Novel Vaccines Against Zika Virus
Published: Jan 2, 2019 Filed: Feb 24, 2017
Earliest Priority: Feb 25 2016 Family: 8 Cited Works: 0
Cited by: 0 Cites: 0 Sequences: 19
Applicant: Univ Pennsylvania

P27 Tyrosine Phosphorylation As A Marker Of Cdk4 Activity

And one can actually go through these particular the basic information provided in relation to that particular sequence. For instance one can get the data in relation to when was the sequence published in relation to that particular application, are there any cited works, is there any patent family information who is the applicant and the basis patent application data. Now one of the interesting aspects of current areas of research. There are many different important and interesting aspects.

(Refer Slide Time: 30:23)



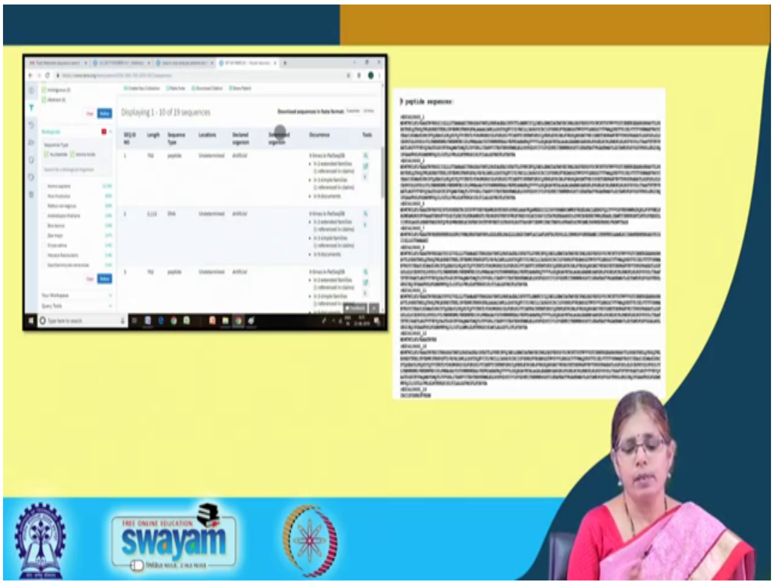
Novel Vaccines Against Zika Virus
Published: Jan 2, 2019 Earliest Priority: Feb 25 2016 Family: 8 Cited Works: 0 Cited by: 0 Cites: 0
Sequences: 19

19 sequences found in this patent (in total)

Filter sequences

If you can see the on this window, you can see the novel vaccines against Zika virus. So, here there are a set of sequences 19 sequences have been submitted by the University of Pennsylvania. So, if you really look at the basic details going from that basic window to the enhanced options that are available, you can actually look at what are all the data available. So, there are nineteen sequences what are the different type of sequences. So, that 10 DNA sequences and 9 peptide sequences and that particular option is available using this particular tool.

(Refer Slide Time: 30:48)



So, one can actually access the peptide sequences. So, in this case you can actually download them in and look at the specific aspects of each of those peptide sequences understand further details in relation to the documents which are submitted for sequences.

(Refer Slide Time: 31:07)

Patent Search with Indian patent office

- http://ipindiaservices.gov.in/publicsearch
- Depending on the status of a patent application, a patent search can be done under two publication types

Patent Search for 18th Month Publications

Public Search for granted Patents

So, these are some of the; these are some of the details that are available in relation to sequence such as available at the databases. One is also aware in the area of biotechnology that there are public databases for the submission of sequence information; for instance EMBL, there is SWISS-PROT all of that, but those will not have the information in relation to the patented sequence information and also the external information.

So, for patented sequence information, one should me to need to go to the aspect of looking at databases that provide for support for patented sequence search. And this is forms a very important part of the search in the case of industries which are based on agriculture biotechnology industries which are into the area of biopharma.

So, the searchers who are from this domain will need to have capacity of looking at the sequence search in relation to patented sequences; patented genetic sequences and patented protein sequences. So, the details of these are available at the WIPO database patent scope and also the LENS dot ORG.