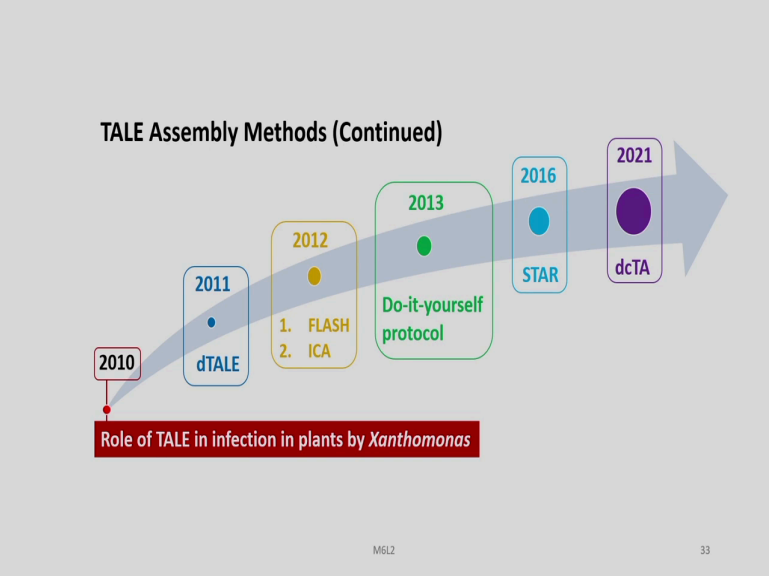
Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 06
Transcription activator-like effector nuclease (TALEN) Technology
Lecture - 02
Design of TALEN for genome editing-PART B

Welcome back to our discussion on the Design of TALENs for genome editing.

(Refer Slide Time: 00:41)

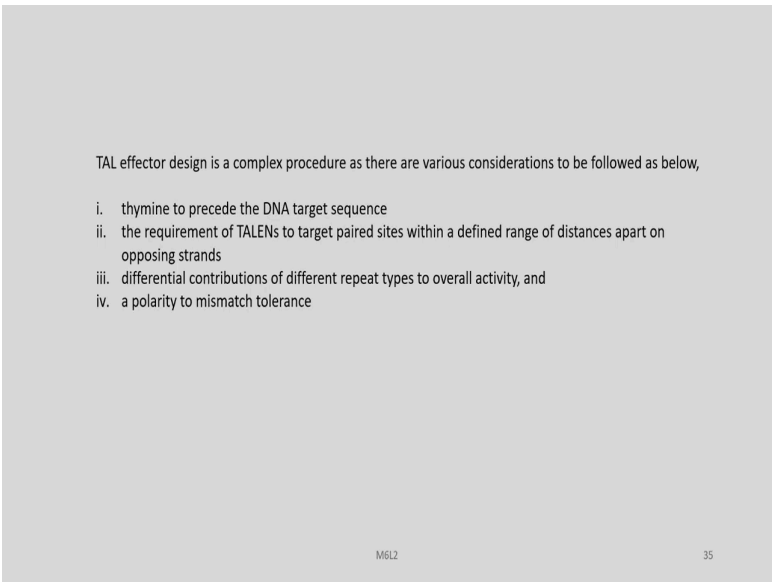


So, in the last part, we discussed about some of the assembly methods which is used for TALE-based fusion protein engineering like dTALE, FLASH, ICA, Do-it-yourself protocol and STAR.

(Refer Slide Time: 00:57)

Today, we will discuss about some of the computational tools that are used in this entire process, as well as for other requirements for TALEN researchers.

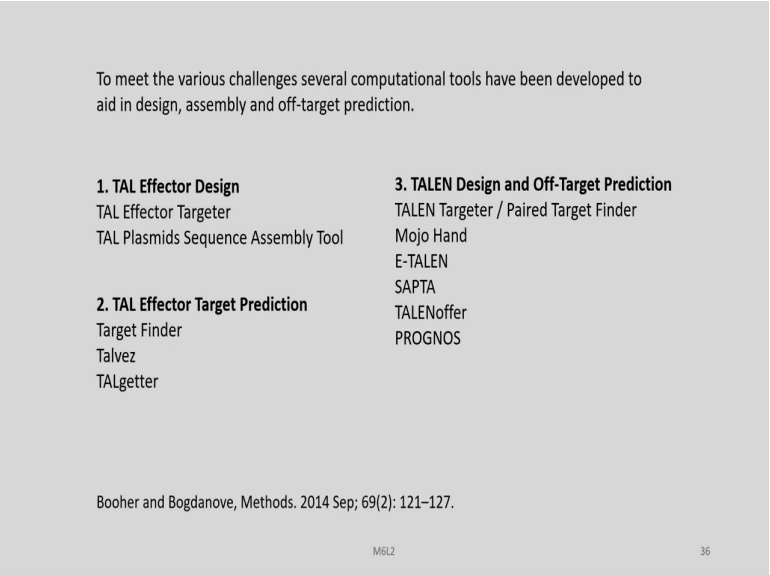
(Refer Slide Time: 01:11)



TAL effector design is a complex procedure as there are various considerations to be followed as below and some of them are already known to you due to the discussions in some of the slides earlier. So, one of the requirements is thymine has to precede the DNA target sequence the requirement of TALENs to target paired sites within defined range of distances

apart on opposing strands. Differential contributions of different repeat types to overall activity and a polarity to mismatch tolerance are required.

(Refer Slide Time: 01:54)

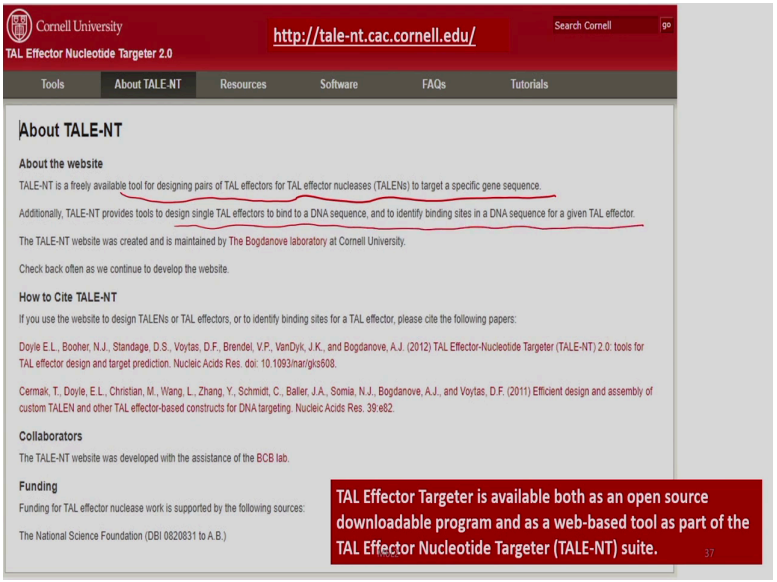


To meet the various challenges as listed above several computational tools have been developed to aid in the design assembly and off-target prediction. So, we can look into these various type of computational tools from three broad divisions: the first division being the software which are being used for TAL effector design examples are TAL effector targeter; TAL plasmids sequence assembly tool.

The second type being TAL effector target prediction. So, here we have software or computational tools likes target finder, then talvez and TALgetter. The third type is the TALEN design and off-target prediction. So, here we have the TALEN targeter/paired target finder, mojo hand, E – TALEN, SAPTA, TALENoffer and PROGNOS.

We will not discuss each and every of these methods, but we will try to discuss at least one or two under each type as listed in these three broad categories.

(Refer Slide Time: 03:13)

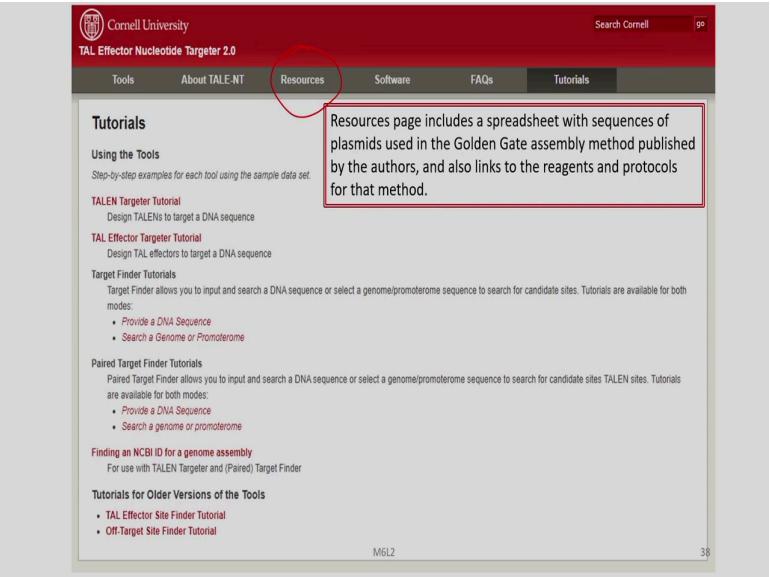


So, about let us go into the discussion of TAL Effector first. So, this is available both as an open software and also as a downloadable program. And, it is a web based interface option is also there and this is part of the TAL effector nucleotide targeter suite, ok, this TALE-NT suite. And, you can log into this by this web link as listed over here.

So, when you log in to this page you will find lot of details about the various software available in this platform and then this one section called resources. So, here you will find the various Resources available under this TALE-NT suite as well as the software part will also list all those various tools available and then if you have certain queries you can go to the FAQs section as usual in all cases of websites.

The best thing is you have tutorials from through which we can learn the usage of these particular tool and these are freely available tools for designing pairs of TAL effectors for TAL effector nucleases to target specific gene sequences. And, additionally, there are many things this provides to design single TAL effectors to bind to a DNA sequence and identify binding sites in a DNA sequence for a given TAL effector.

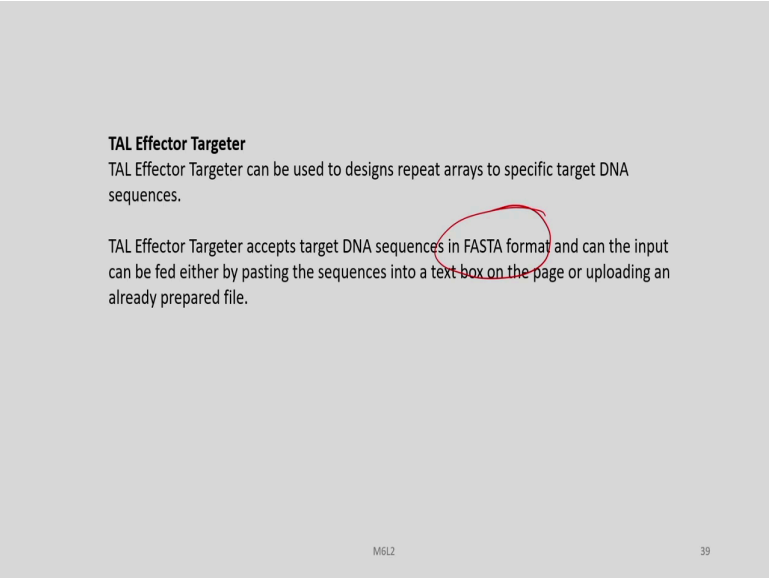
(Refer Slide Time: 05:02)



So, this is the tutorial site I was speaking about: how to use these tools. There are different tools here: the TALEN Targeter Tutorial, the TAL Effector Targeter Tutorial and so on and so forth. And, then you have the tutorials for older versions of the tools. So, if you have some of the old software downloaded then you can get the tutorials in this particular linkage.

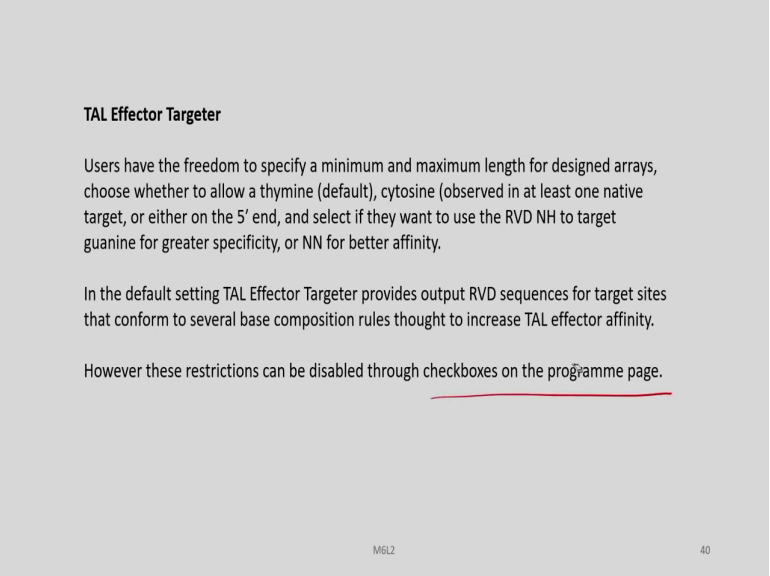
This Resource page includes a spreadsheet with sequences of plasmids used in the golden gate assembly method published by the authors and also links to the reagents and protocols for that particular method.

(Refer Slide Time: 05:54)



The TAL Effector Targeter can be used to design repeat arrays to specific target DNA sequences. The TAL effector targeter accepts target DNA sequences in FASTA format as in many bioinformatics tools and software or web interfaces and the input can be fed either by pasting the sequence into a text box on the page or uploading an already prepared file.

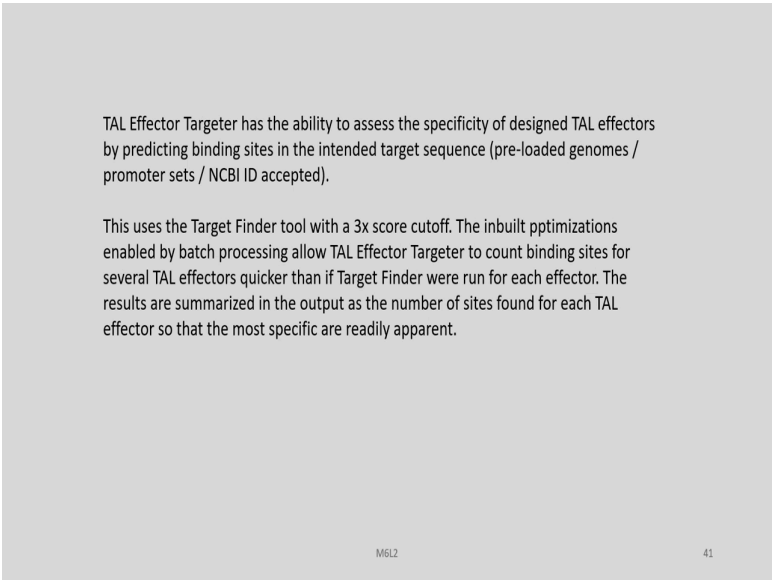
(Refer Slide Time: 06:24)



In this particular platform, the users have the freedom to specify a minimum and maximum length for designed arrays, and they can choose whether to allow thymine (default) or cytosine (observed in at least one native target), or either on the 5 prime end and select if they want to use the RVD NH to target guanine or for greater specificity or NN for better affinity.

In the default settings TAL Effector Targeter provides output RVD sequences for target sites that conform to several base composition rules throughout to increase TAL effector affinity. However, these restrictions can be disabled through checkboxes on the programme page.

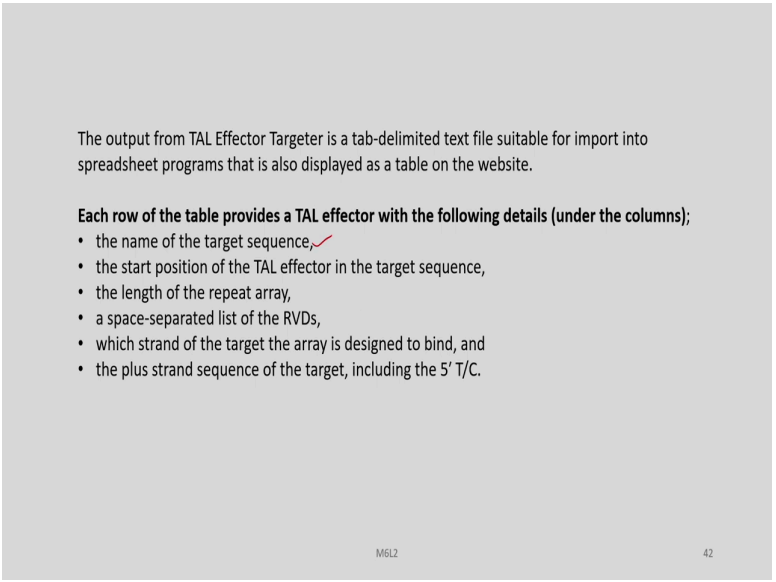
(Refer Slide Time: 07:19)



The TAL Effector Targeter has the ability to assess the specificity of design TAL effectors by predicting binding sites in the intended targeted sequence the preloaded genomes promoter sets NCBI IDs are accepted.

This uses the Target Finder tool with a 3x score cutoff. The inbuilt optimizations enabled by batch processing allows TAL Effector Targeter to count binding sites for several TAL Effector quicker than if TAL Finder were run for each effector. The results are summarized in the output as the number of sites found for each TAL effector so that the most specific ones are readily apparent.

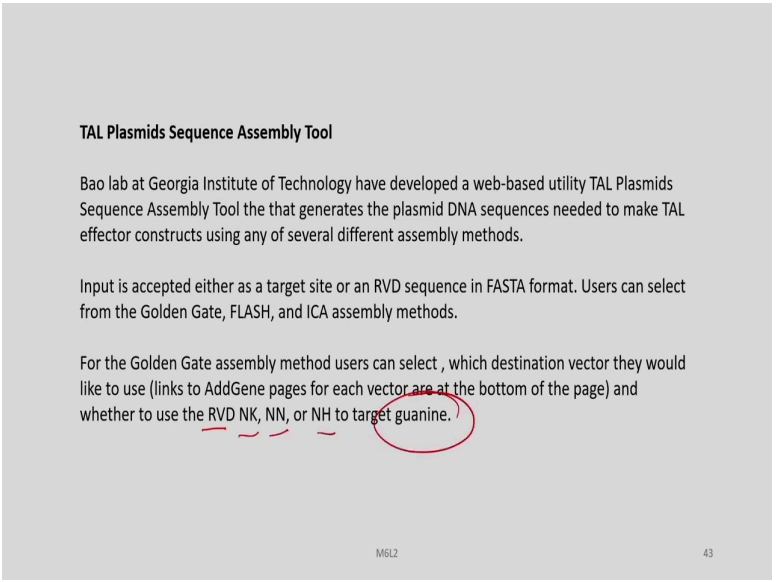
(Refer Slide Time: 08:05)



The output from TAL Effector Targeter is a tab-delimited text file which is suitable for import into spreadsheet programs that is also displayed as a table on the website.

So, each row of the table provides a TAL effector with the following details under columns. It will give the name of the target sequence, the start position of the TAL effector in the target sequence, the length of the repeat array, a space-separated list of the RVDs, which strand of the target the array is designed to bind, and the plus strand sequence of the target including the 5 prime T/C.

(Refer Slide Time: 08:48)

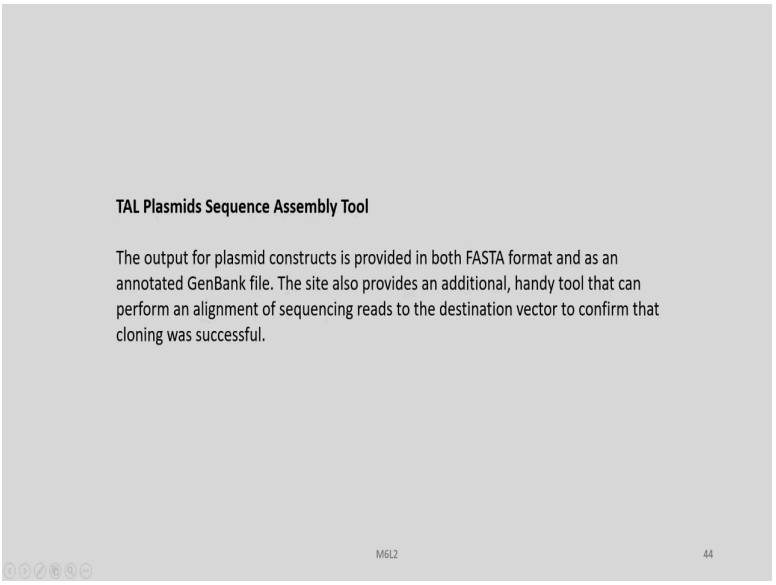


Let us now discuss about the TAL plasmids sequence assembly tool. Bao lab at Georgia Institute Technology developed this web-based utility called TAL Plasmids Sequence Assembly Tool. It generates the plasmid sequences needed to make TAL effector constructs using any of the several different assembly methods.

Input is accepted either as a target site or an RVD sequence in FASTA format. Users can select from the Golden Gate, FLASH, and ICA assembly methods. So, we have discussed about these methods in detail or briefly in the earlier part.

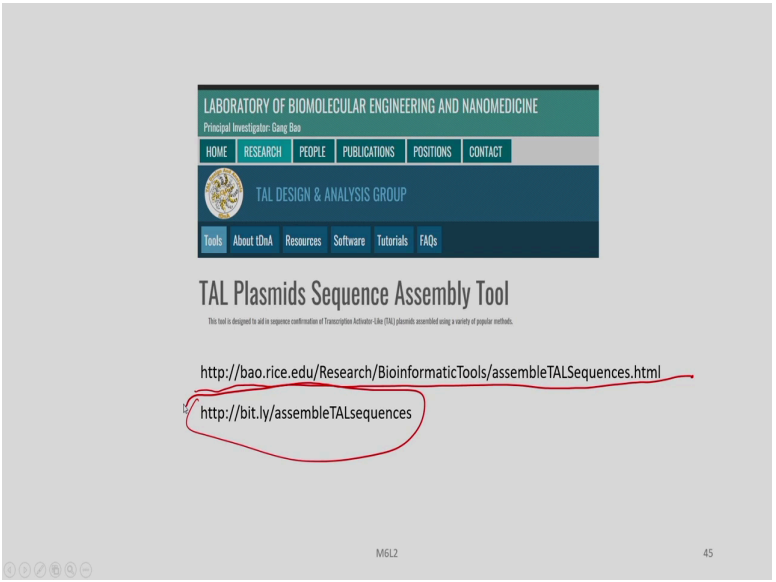
For the Golden Gate assembly method users can select, which destination vector you would use like to use and here you have links to the AddGene pages for each vector at the bottom of the page and whether to use the RVD NK, NN or NH to target guanine.

(Refer Slide Time: 09:59)



The output for plasmid construct is provided in both FASTA format and as an annotated GenBank file. The site also provides an additional, handy tool that can perform an alignment of sequencing reads to the destination vector to confirm that cloning was successful.

(Refer Slide Time: 10:19)

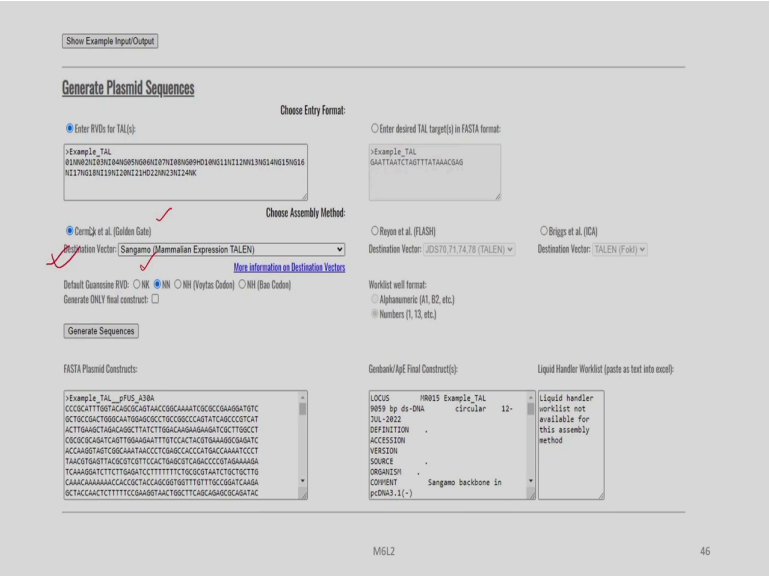


So, this is the laboratory of biomolecular engineering and nanomedicine and you have here Gang Bao is the principal investigator. This is the TAL Design and Analysis Group you can visit this site. You can get into the resources site to know what else is available as well as the software and then the tutorial part through which you can learn many of the applications and various fundamental basic questions are already answered in this FAQs.

And this is the website:

http://bao.rice.edu/Research/BioinformaticsTools/assembleTALsequences.html or there is a short bit.ly address as well, through which you can log in to these particular websites.

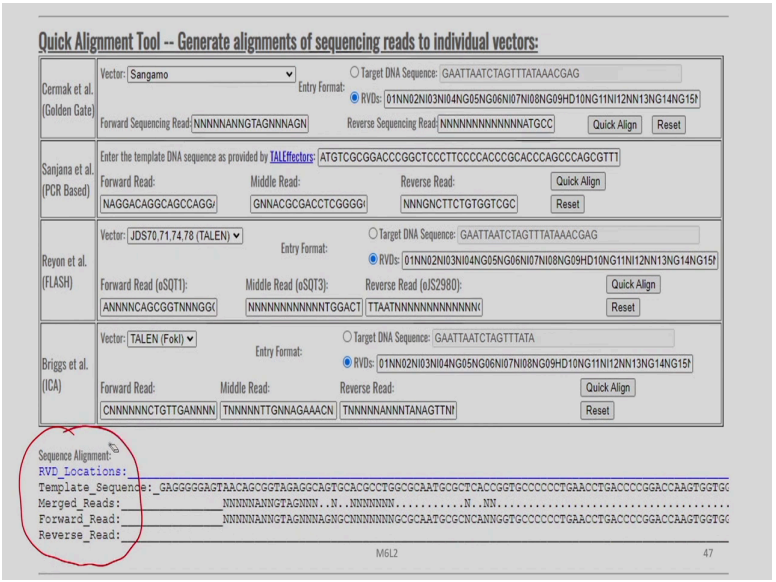
(Refer Slide Time: 11:15)



So, let us show some examples of these particular application to generate plasmid sequences. So, you can enter RVDs for the TALs here or you can enter desired TAL targets in FASTA format ok.

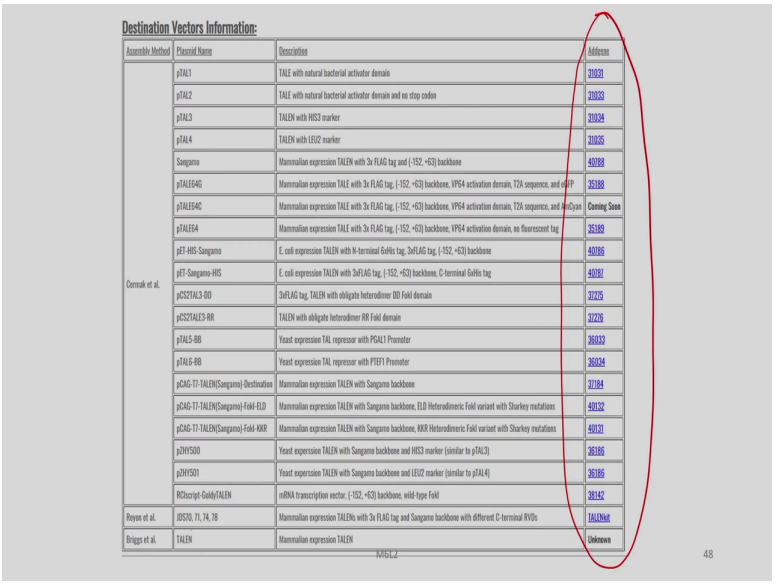
Then you have the assembly method which you can choose. You can use Golden Gate, you can use FLASH, you can use ICA. And then if you click into this drop down menu you can get many of the destination vectors which you can use for your purpose. And, then you can use these options for the Default Guanosine RVD, either of these. So, in this case the NN has been used and then you can generate the sequences and you can see here the FASTA plasmid construct sequence.

(Refer Slide Time: 12:29)



So, anyway this is quite simple. So, this has a quick alignment tool which generate alignments of sequencing reads to individual vectors as well and you can see here the Sequence Alignment taking place.

(Refer Slide Time: 12:58)

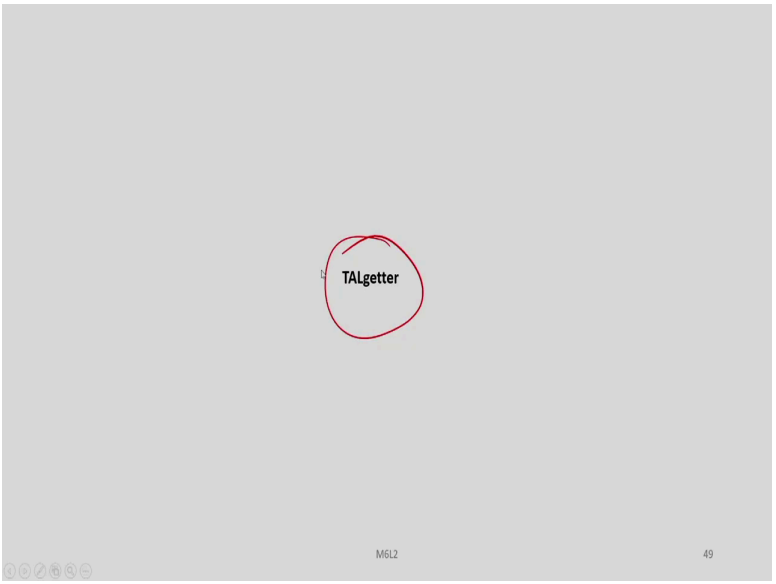


And, we have various options which you can choose as per your requirement in this entire design. So, the Destination Vectors which you can see here, right. This is the list of the Destination Vectors at your disposal under this software, so many of them.

So, you have the description over here for example, pTAL1 is a TAL effector with natural bacterial activator domain and pTAL2 is a natural pTAL bacterial activator domain with no stop codon. Then you have certain plasmids like Sangamo which is a mammalian expression TALEN with thrice x FLAG tag and you have many such vectors which are useful for Mammalian expression and you can use them for animal experimentation.

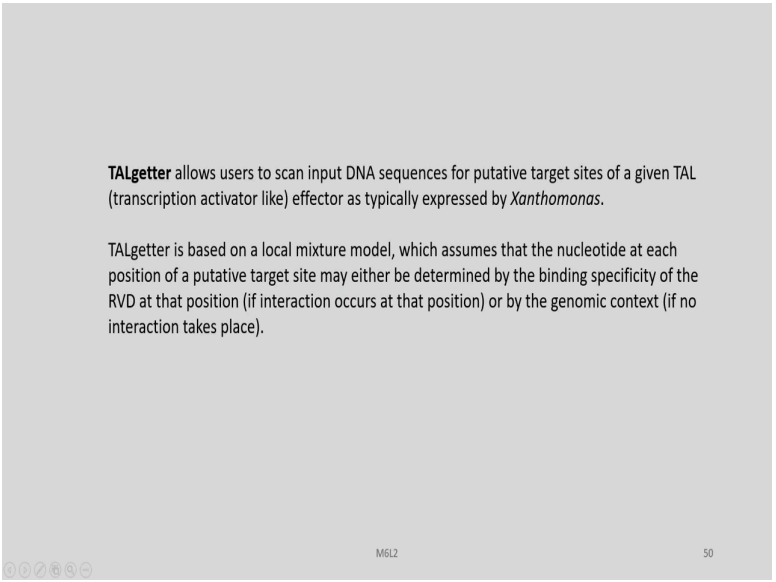
Then even Briggs et al. this is the plasmid name as TALEN and this is the Mammalian expression TALEN. And, this is the link of the Addgene repository. So, if you click into these blue highlighted links, you can land up in a particular page and you can find out the details, availability and how to order them.

(Refer Slide Time: 14:36)



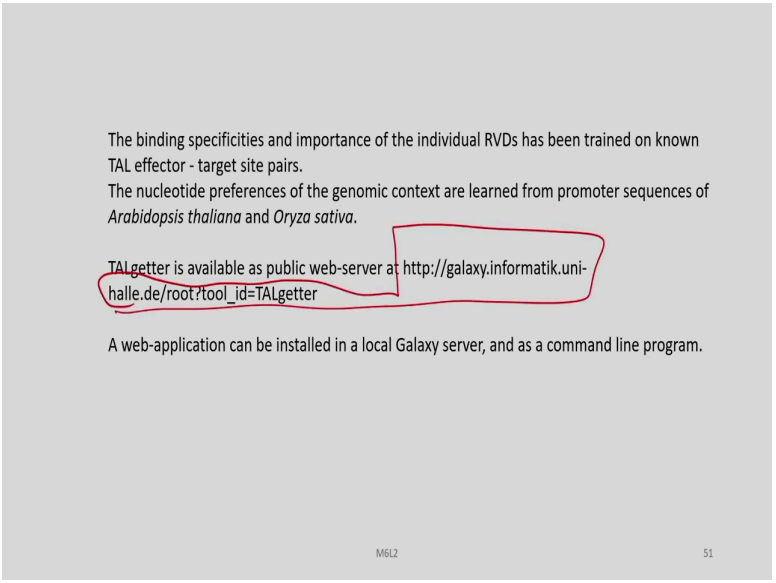
Let us now discuss another such tool which is the TALgetter.

(Refer Slide Time: 14:44)



The TALgetter allows users to scan input DNA sequences for putative target sites of a given TAL effector as typically expressed by *Xanthomonas*. TALgetter is based on a local mixture model, which assumes that the nucleotide at each position of a putative target site may either be determined by the binding specificity of the RVD at that position or if interaction occurs at that position or by the genomic context if no interaction takes place.

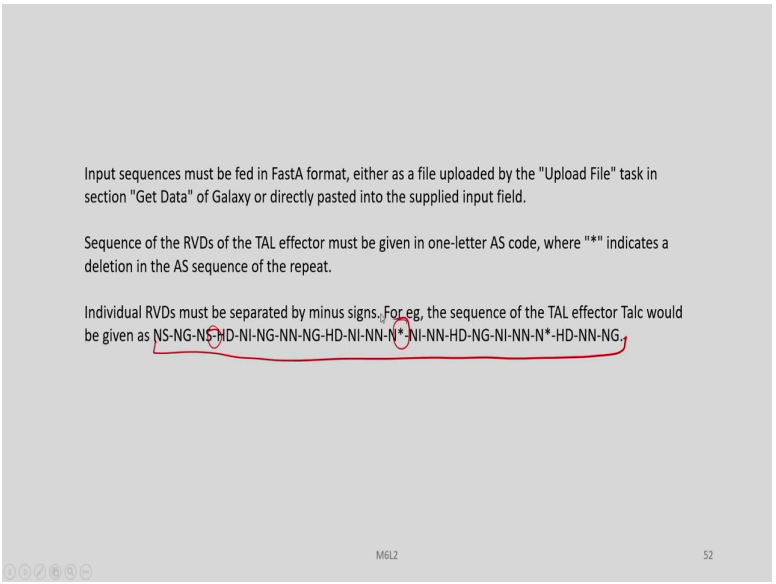
(Refer Slide Time: 15:20)



The binding specificities and importance of the individual RVDs has been trained on known TAL effector – target site pairs. The nucleotide preferences of the genomic context are

learned from promoter sequences of *Arabidopsis thaliana* and *Oryza sativa*. TALgetter is available as a public web-server in this web link. A web-application can be installed in the local server Galaxy server and as a command line program as well.

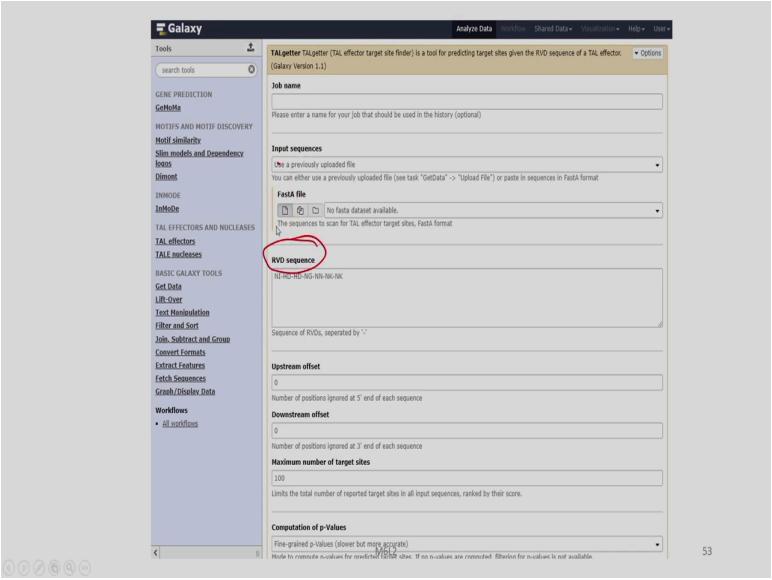
(Refer Slide Time: 16:57)



And, as in the case of others the input sequences must be fed in FASTA format which is a standard and either it can be uploaded as a file or in the task sections Get Data of Galaxy or directly pasted into the supplied input field. Sequence of the RVDs of the TAL effectors must be given in one-letter AS codes, where these star asterisks indicates a deletion in the AS sequence of the repeat.

Individual RVDs must be separated by minus signs. For example, the sequence of the TAL effector Talc would be given as NS then this is the minus sign NG minus sign NS and so on and then this star shows a deletion, ok. So, these are very important points to remember to get the tool working for us in designing effective TAL effectors.

(Refer Slide Time: 17:11)

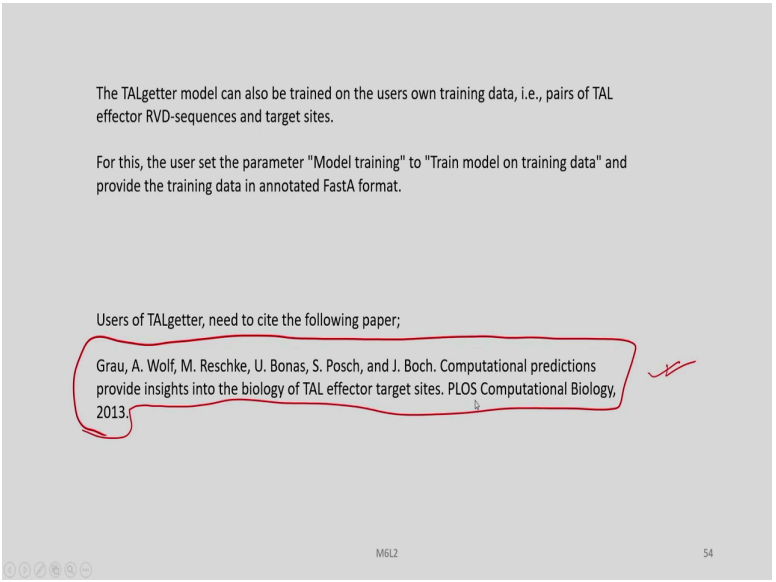


So, here as we were telling that you have the Galaxy website where you work under the section Get Data. So, you can give the Job name here, ok. You just name it as per your convenience then this is the input sequence where you can upload these in the FASTA file format and then or you can paste it also as well.

And, these RVD sequence which we are referring to here including the minus signs as well as the star marks or asterisks marks. This has to be defined over here. So, it is written of course. Sequence of RVDs separated by the dash or minus sign. So, you can have Upstream offset, you can have Downstream offset and Maximum number of target sites as well you can define and some Computational p-Values can also be defined.

So, you can visit this by visiting these website and practice on your own taking any of the targets of your choice and design, some of the TAL domains.

(Refer Slide Time: 18:46)

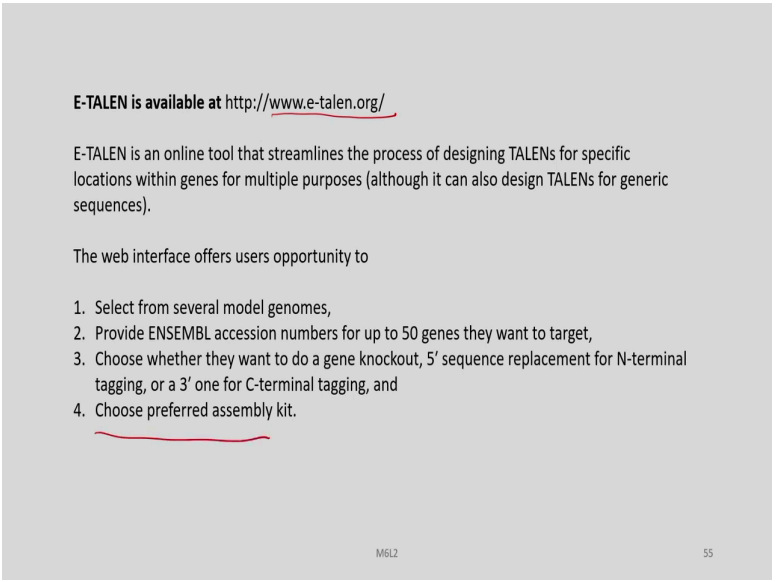


These TALgetter model can also be trained on users own training data. So, this has been originally trained with data from *Arabidopsis thaliana* and *Oryza sativa* as I have told you earlier, but you can also train it on your own data that is pairs of TAL effector RVD sequences and the target sites.

But, to do this as a user you have to set the parameter “Model training” change it to “Train model on training data” and provide the training data in annotated FASTA format, ok. So, changing the set parameter from “Model training” to “Train model on training data” is important for training on the user’s data and which has to be provided as annotated FASTA format.

And, if you are a user of this particular tool it is being requested to site the developers in this citation as I have listed over here and published in a PLOS Computational Biology in the year 2013.

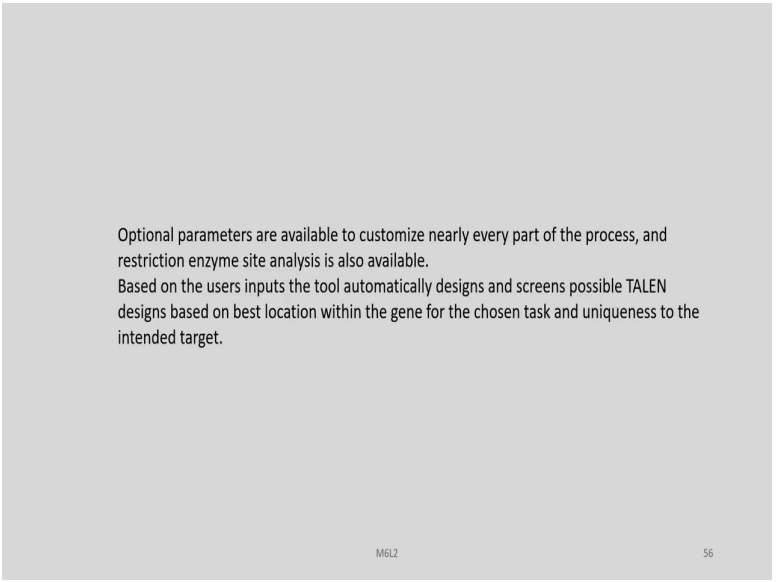
(Refer Slide Time: 20:20)



So, another tool that is available is E-TALEN which is available in the website https://www.e-talen.org. This is an online tool which streamlines the process of designing TALENs for specific locations within genes for multiple purposes although it can also design TALENs for generic sequences.

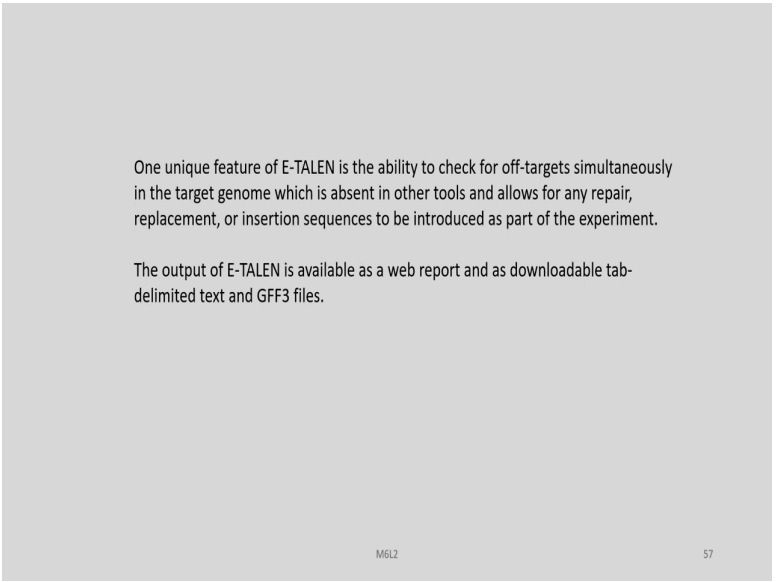
The web interface offers users various opportunities like selecting from several model genomes; provide ENSEMBL accession numbers for up to 50 genes they want to target; choose whether they want to do a gene knockout, 5 prime sequence replacement for N-terminal tagging or a 3 prime one for C-terminal tagging and choose a preferred assembly kit.

(Refer Slide Time: 21:16)



Optimization parameters are available to customize nearly every part of the process and restriction enzyme site analysis is also available. Based on the users input the tool automatically designs and screens possible TALEN designs based on best location within the gene for the chosen task and uniqueness to the intended target.

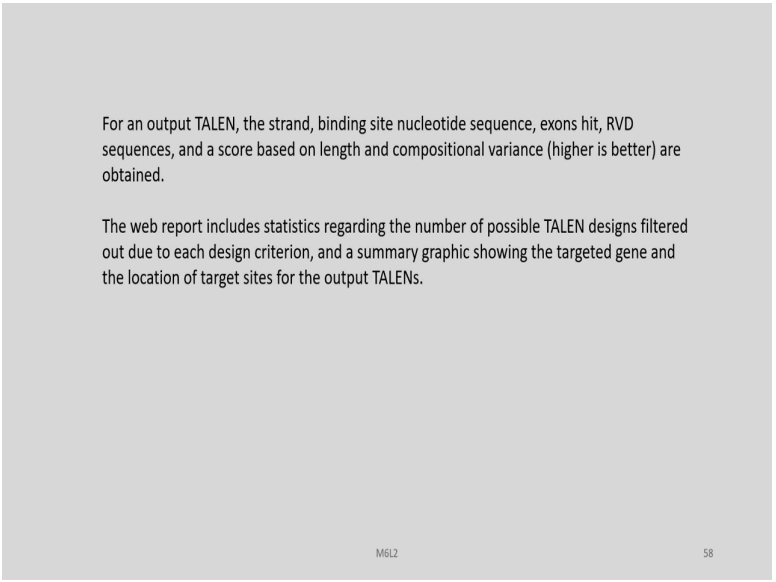
(Refer Slide Time: 21:41)



One unique feature of E-TALEN is the ability to check for off-targets simultaneously in the target genome which is absent in other tools and allows for any repair, replacement, or insertion sequences to be introduced as part of the experiment.

The output of E-TALEN is available as a web report and as downloadable tab-delimited text and GFF3 files.

(Refer Slide Time: 22:11)



For an output TALEN, the strand, binding site nucleotide sequence, exons hit, RVD sequences, and a score based on length and compositional variance, if the higher value is more it is better, are obtained.

The web report includes also statistics regarding the number of possible TALEN designs filtered out due to its design criterion, and a summary graphic showing the targeted gene and the location of target sites for the output TALENs and these are not these are unique to this particular program.

So, thank you for your patient hearing.