

Tutorial

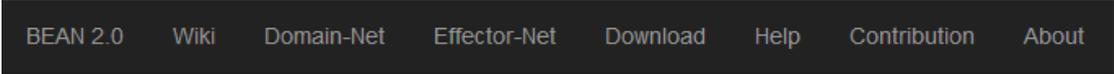
This tutorial introduces how to use this web server to predict and analyze possible type-III effectors. The web server allows anonymous use. If you want to keep your query sequences confidential and manage your jobs, please register and login using username.

This website can be divided into three parts.

The first part is BEAN 2.0. BEAN 2.0 uses features learned from known type-III effectors to predict query proteins as effectors or non-effectors.

The second part is designed for the further analysis of proteins, including four components. i) By searching a protein sequence against the Pfam library of HMMs, you can determine which domains it carries. ii) TargetP predicts the subcellular location of eukaryotic proteins. iii) Plant-mPloc and Hum-mPloc predict subcellular localizations of plant and non-plant proteins including those with multiple sites, respectively. v)IUpred is used to predict the likelihood of disorder for each residues.

At last, we present a specialized database of annotated type-III effectors.



BEAN 2.0 Wiki Domain-Net Effector-Net Download Help Contribution About

In the navigation bar, there are also some other conclusive information. For example, the nets of the domains and effectors are showed in Domain-net and Effector-net, respectively. Users can check the details of the networks just by spinning the wheels of the mouse. Users are also allowed to target at only one network by put the mouse on the node of this network.

In addition, you may click "Contribution" to make suggestions and provide effectors. We are very grateful for your suggestions and sharing. If you have some other questions, please don't hesitate to contact me.

FASTA format

BEAN 2.0 uses sequences in FASTA format to predict possible type-III effectors. FASTA is one of the most common formats that are used to represent DNA or protein sequences in modern biology. A FASTA sequence. A FASTA sequence should be presented as:

```
>IPAD_SHIFL
MNIITLTLNSISTSSFSFNNTNGSSTETVNSDIKTTTSSHPVSSLTMLNDTLHNIIRTTNQA
LKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRNEYYPINKDARELLHSAPKE
AELDGDQMI SHRELWAKIANSINDINEQYLKVYEHAVSSYTQMYQDFSAVLSSLAGWISP
GGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTVSEQANKWLTELGGTIGKVSQK
NGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKYQAWNAGFSAEDETMMKNNLQTLVQK
YSNANSIFDNLVKVLSSTISSCTDTDKFLHF
```

The first line starts with a ">", and follow with the sequence name. The other lines are the protein sequence represented in single letter form. If you have more than one sequence, you just need to write each of them one by one in the same way (Please don't leave any blank lines between them!).

BEAN 2.0

Using webserver

Prediction

You need to provide full-length amino acid sequence(s) in *fasta* format of query protein(s) using BEAN. The maximal number of sequences you can submit to our webserver is 50 per job.

We will send an e-mail to your address when the job is finished. [Example](#)

Filename:

No file selected.

Job Name*

Email Address*

Click "start new job" to predict candidate proteins. You need to input your protein sequences in FASTA format into the textfield. Then please input your email address and submit. Each sequence may take 2 minutes or more. When your job is finished, you will receive an email. BEAN 2.0 will return the result reporting whether the query proteins are effectors or not in the webpage that is included in the email. Then you can download the results or further analyze the corresponding domains, subcellular locations and disorder

residues of these proteins.

You can also check submitted jobs. You can login the system and manage your jobs if you have a account.

Running BEAN 2.0 on your local machine

Please refer to the installation guide wrapped in BEAN 2.0's software package or turn to download webpage.

Output

Select All [Download](#) [Analysis](#) Method 1: sequence alignment-based prediction 2: domain-based prediction 3: machine-learning

Protein	Score/E-value	Method	Is type III effector?	Select
sp Q72HB4 YF81_THET2	1e-25	1	no	<input type="checkbox"/>
tr G8Z8Z9 G8Z8Z9_BRAOL	-0.616122	3	no	<input type="checkbox"/>
sp Q9FJA2 TT2_ARATH	-0.141079	3	no	<input type="checkbox"/>
sp Q39258 VATE1_ARATH	-0.840769	3	no	<input type="checkbox"/>
sp Q9LFC9 E2FD_ARATH	-0.0647995	3	no	<input type="checkbox"/>
sp Q66K89 E4F1_HUMAN	-1.26494	3	no	<input type="checkbox"/>

Algorithm

Click "wiki" to see the algorithm of BEAN 2.0.

Methods

BEAN 2.0 is a hybrid approach combined of BLAST search, Pfam domain and BEAN using features of three parts of protein sequences. If the candidate protein is predicted by BLAST or Pfam, e-value will be showed to indicate the confidence level of the result. Otherwise, raw score will be used to classify a unknown protein sequence as candidate type-III effectors or not.

Raw Score

BEAN* uses the raw score to classify a unknown protein sequence as candidate type-III effectors or not. The higher the score is, the more possible the protein is a type-III effector. BEAN 2.0 uses score=0.0 as a cutoff to make the decision, whereby all proteins

assigned with scores equal or above 0.0 are considered as candidate TTEs.

E-value

BLAST and Pfam use E-value as a threshold to judge whether the candidate proteins can be predicted as effectors. The Expect value used as a convenient way to create a significance threshold for reporting results. The lower the E-value, the more significant the match is. The closer it is to zero, the results tend to be more reliable.

Analysis

Using webserver

Analysis

You need to provide full-length amino acid sequence(s) in [fasta](#) format of query protein(s). The maximal number of sequences you can submit to our webserver is 50 for per job. We will send an e-mail to your address when the job is finished. [example](#)

Filename:

未选择文件

Job name*

Predict Subcellular Location: Plant TargetP(plant) Plant-mPLoc Non-plant TargetP(non-plant) Hum-mPLoc

Scan Pfam Domain: Pfam

Predict Protein Disorder: IUPred

Click "analysis" to predict the domains, subcellular locations and disorder residues of query proteins. You need to input your protein sequences in FASTA format into the textfield, and give a name consisted with alphanumeric characters or underscores to demark you job. Then you also need to choose a type that you want to predict. After a few minutes ,you will get the result. You can also download your selected analysis results, too.

Tools

Pfam

The Pfam database is a large collection of protein domain families. Each family is represented by multiple sequence alignments and hidden Markov models (HMMs). By searching a protein sequence against the Pfam library of HMMs, you can determine which domains it carries.

TargetP

TargetP predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP).

For the sequences predicted to contain an N-terminal presequence a potential cleavage site can also be predicted.

Plant-mPloc

Plant-mPloc predict subcellular localizations of plant proteins including those with multiple sites.

Hum-mPloc

Hum-mPloc predict subcellular localizations of non-plant proteins including those with multiple sites.

IUPred

IUPred is a predictor of prediction of intrinsically unstructured proteins. Intrinsic disorder is particularly enriched in proteins implicated in some important functions.

Output

Pfam

Select All

Name	Envelope Start	Envelope End	Hmm Acc	Hmm Name	Type	Hmm Length	Bit Score	E-value	Significance	Clan	Select
OPGH_SALTY.pssm	245	430	PF00535.21	Glycos_transf_2	Family	169	66.6	2e-18	1	CL0110	<input type="checkbox"/>
OPGH_SALTY.pssm	535	622	PB000091	Pfam-B_91	Pfam-B	147	68.6	5.5e-19	NA	NA	<input type="checkbox"/>
OPGH_SALTY.pssm	622	765	PB002938	Pfam-B_2938	Pfam-B	445	49.5	4.1e-13	NA	NA	<input type="checkbox"/>



Name	Envelope Start	Envelope End	Hmm Acc	Hmm Name	Type	Hmm Length	Bit Score	E-value	Significance	Clan	Select
RL7_CHLTR.pssm	60	128	PF00542.14	Ribosomal_L12	Domain	68	83.2	9.4e-24	1	No_clan	<input type="checkbox"/>



In addition to the results, you will get a graphical interface that displays the domains on the query proteins. For more information can be found at <http://pfam.xfam.org/help>.

TargetP

Select All

Name	Len	CTP	MTP	SP	Other	Loc	RC	TPlen	Select
H6UWS2_XANOR	137	0.282	0.322	0.002	0.200	M	5	64	<input type="checkbox"/>

Name	Sequence name truncated to 20 characters
Len	Sequence length
cTP,mTP,SP,other	Final NN scores on which the final prediction is based (Loc, see below). Note that the scores are not really probabilities, and they do not necessarily add to one. However, the location with the highest score is the most likely according to TargetP, and the relationship between the scores (the reliability class, see below) may be an indication of how certain the prediction is.
Loc	Prediction of localization, based on the scores above; the possible values are: C Chloroplast M Mitochondrion S Secretory pathway

	<p>_ Any other location</p> <p>* "don't know"</p>
RC	<p>Reliability class, from 1 to 5, where 1 indicates the strongest prediction. RC is a measure of the size of the difference ('diff') between the highest(winning) and the second highest output scores. There are 5 reliability classes, defined as follows:</p> <p>1 : $\text{diff} > 0.800$</p> <p>2 : $0.800 > \text{diff} > 0.600$</p> <p>3 : $0.600 > \text{diff} > 0.400$</p> <p>4 : $0.400 > \text{diff} > 0.200$</p> <p>5 : $0.200 > \text{diff}$</p> <p>Thus, the lower the value of RC the safer the prediction.</p>
TPlen	<p>Predicted presequence length; it appears only when TargetP was asked to perform cleavage site predictions</p>

For more information, please visit <http://www.cbs.dtu.dk/services/TargetP/output.php>.

Plant-mPloc

<input type="checkbox"/> Select All	Download	
Query Protein	Predicted Location(s)	Select
H6UWS2_XANOR	Nucleus	<input type="checkbox"/>

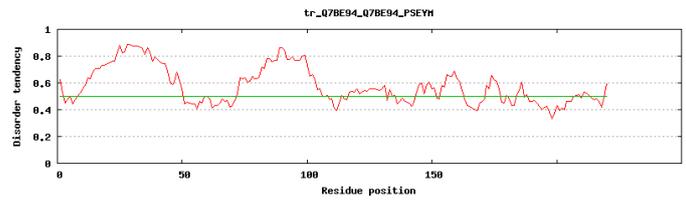
You will get a table contain the subcellular locations of the query proteins. Due to the instability of this predictor, the server will automatically turn to TargetP if there were no results after 100s.

Hum-mPloc

<input type="checkbox"/> Select All	Download	
Query protein	Predicted location(s)	Select
H6UWS2_XANOR	Cytoplasm	<input type="checkbox"/>

You will get a table contain the subcellular locations of the query proteins. Due to the instability of this predictor, the server will automatically turn to TargetP if there were no results after 100s.

IUPred



```
# tr_Q7BE94_Q7BE94_PSEYM
MGCVSSTSRSTGYSCYENHIEPRVASSPTNRHRDGYETDSEHSNVNLIPNARRVYKESLWHGTSMQSKIDLRRGQFDVNRKTDGATEGSRSTSNSAPTDLTRNARRHNYLTAYDYTKRNYARRADPEPAPAVRTIGVKSNNF
TELDPEKRDENGETSQSCRTTDSIPKRYLVGSKSQPGADAKWFKKELSNAGREVSISRAGELRRAVQSDSEDDI
```

The output gives the likelihood of disorder for each residue, i.e. it is a value between 0 and 1, and higher values indicate higher probability of disorder. Residues with value above 0.5 can be regarded as disordered. The webserver showed the value use a rainbow bar. The color change from warm to cool with the increasing of the value.

Database

Type III Secreted Effector Database											download	
All (1)											All (21)	All (226)
Experimentally validated effector											<ul style="list-style-type: none"> Animals Animals and Humans Aras Guinea pigs Guinea pigs 	<ul style="list-style-type: none"> Acidovorax avenae subsp. avenae Aeromonas hydrophila Aeromonas salmonicida Aeromonas salmonicida (strain A449) Aeromonas veronii
UniprotID	Name	Function	Host	Pathogen	Source	Length	Domain	Subcellular location	TargetP	inPloc		
Q63K50_PURE3	Uncharacterized protein Q63K50	Unknown	Animals and Humans	Burkholderia pseudomallei (strain 306243)	Experimentally validated effector	469	Phosphorylase superfamily	Unknown	Any other location	Cytoplasm		
Q9TGF9_Y1BFA	Uncharacterized protein Q9TGF9	Unknown	Shellfishes	Vibrio parahaemolyticus serotype O3:K6 (strain NMD 2210633)	Experimentally validated effector	1622	Apolipoprotein L	Unknown	Any other location	Nucleus		
Q7RE90_SHIFL	Enterotoxin	Unknown	Animals and Humans	Shigella flexneri 5a	Experimentally validated effector	565	ShET2 enterotoxin, N-terminal region	Unknown	Any other location	Cytoplasm		
IPAA_SHIFL	Invasin IpaA	Rapidly associates with the first 265 amino acids of vinculin after bacterium-cell contact. This interaction is critical for efficient Shigella uptake. IpaA acts as a potent activator of vinculin and increase its ability to interact with F-actin. The complex IpaA-vinculin induces F-actin depolymerization along with the occasional formation of actin filament bundles.	Animals and Humans	Shigella flexneri	Experimentally validated effector	633	Salmonella invasion protein A	Cytosol	Any other location	Nucleus		
IPAB_SHIFL	Invasin IpaB	Effector proteins function to alter host cell physiology and promote bacterial survival	Animals and Humans	Shigella flexneri	Experimentally validated effector	580	Secretion system effector C (SscC) like family	Membrane;Nucleus	Any other location	Cytoplasm		

This database contains effector records. We collected these records mainly from Uniprot. Search results are presented in a tabular form, displaying effector name, source organism of the effector, sequence length, experimental status, domain(Pfam), subcellular locations in host. You can also download the database.

Exclusive domains in Type-III effectors have been summarized and displayed in "Domain" page.