

Supporting Information for:

motifeR: An integrate web server for identification of protein post-translational modification motifs and inference of kinase activity

Shisheng Wang,[†] Yue Cai,[‡] Jingqiu Cheng,[†] Yansheng Liu,^{*,§} and Hao Yang^{*,†}

[†] West China-Washington Mitochondria and Metabolism Research Center; Key Lab of Transplant Engineering and Immunology, MOH, West China Hospital, Sichuan University, Chengdu, China.

[‡] Ministry of Science and Technology, West China Hospital, Sichuan University, Chengdu, China.

[§] Department of Pharmacology, Yale University School of Medicine, New Haven, CT 06520, USA.

Corresponding Author

*Email address: yansheng.liu@yale.edu; yanghao@scu.edu.cn.

Table of Contents

1. Supplementary Figures and Tables.

Figure S1. The current status of some published tools.

Figure S2. Distribution of peptide modification site number.

Table S1. The example of pre-alignment results. (Please see the online file).

Table S2. The example of kinase-substrate network analysis. (Please see the online file).

2. Supplementary notes.

2.1 Data Preparation.

2.2 Import data.

2.3 Pre-alignment.

2.4 Own Background.

2.5 Motif Enrichment.

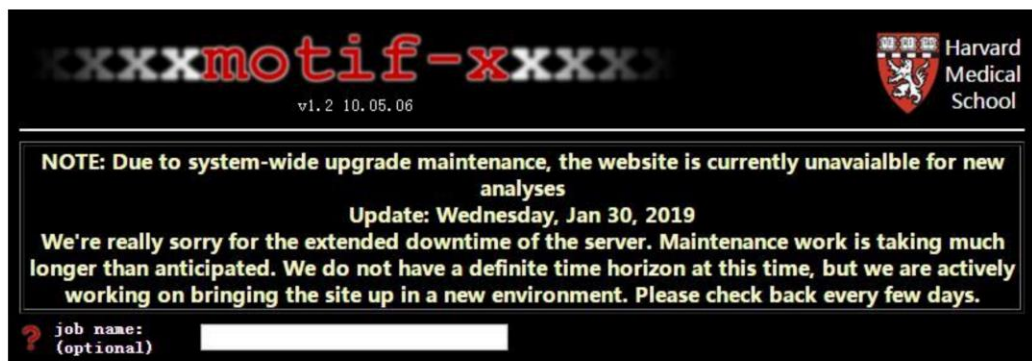
2.6 Motif Plot.

2.7 Kinase-Substrate Analysis.

2.8 Building Species Database.

Supplementary Figures

A



B

Discovery of biological sequence motifs in R

62 commits

1 branch

0 releases

3 contributors

Branch: master

New pull request

Find File

Clone or download

omarwagih Update README.md

Latest commit #13775f on Dec 20, 2018

R	added citation	4 years ago
build	added citation	4 years ago
inst/extdata	added citation	4 years ago
man	package name changed	a year ago
.gitignore	DS_store banished!	5 years ago
DESCRIPTION	update name to rmotifx	5 years ago
NAMESPACE	first commit	5 years ago
README.md	Update README.md	4 months ago

Figure S1. The current status of some published tools. A. motif-x has been in maintenance for a while and users don't know when it can be back. (Link: <http://motif-x.med.harvard.edu/motif-x.html>). B. rmotifx package has not been updated in a long time. (Link: <https://github.com/omarwagih/rmotifx/>).

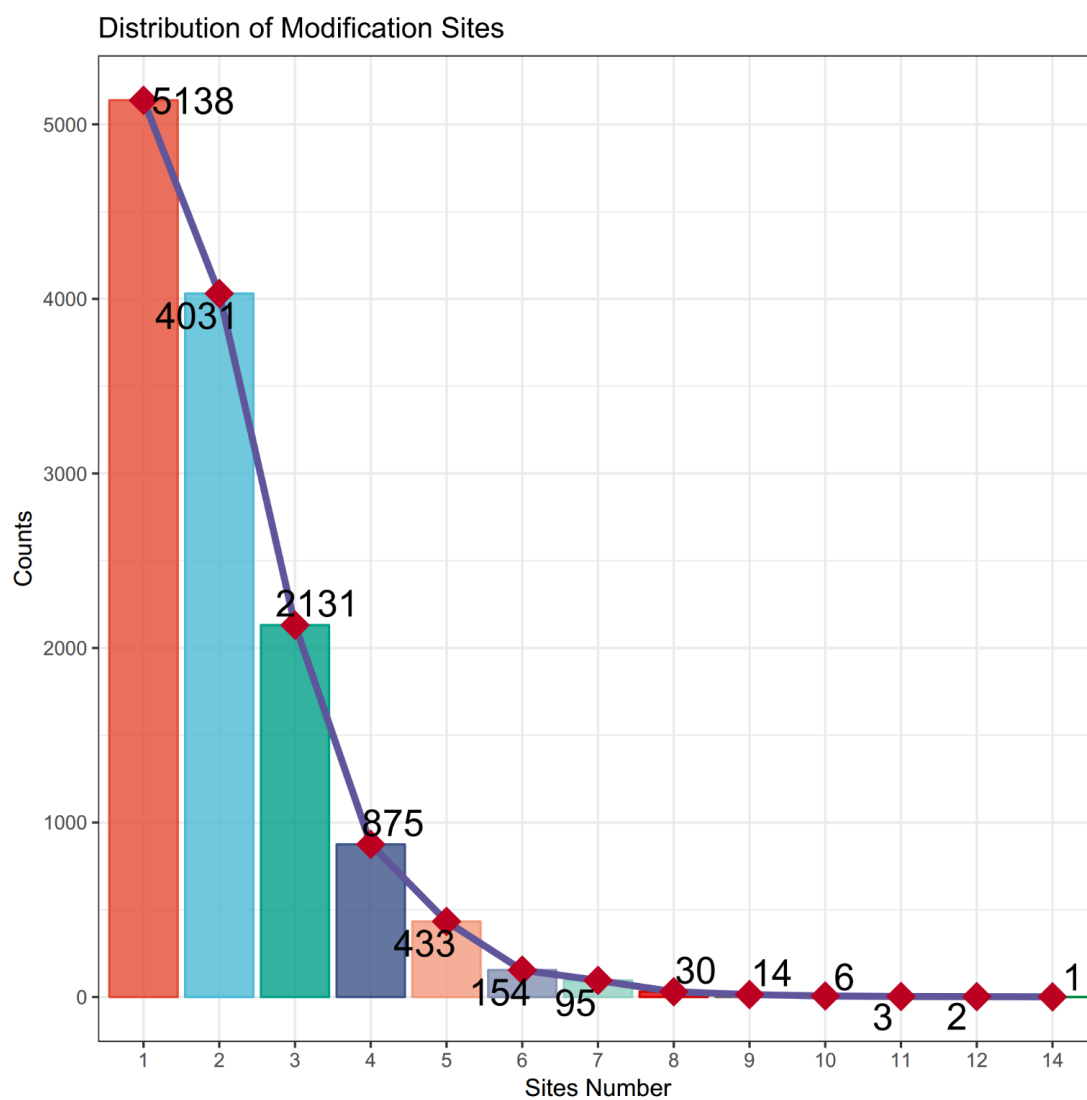
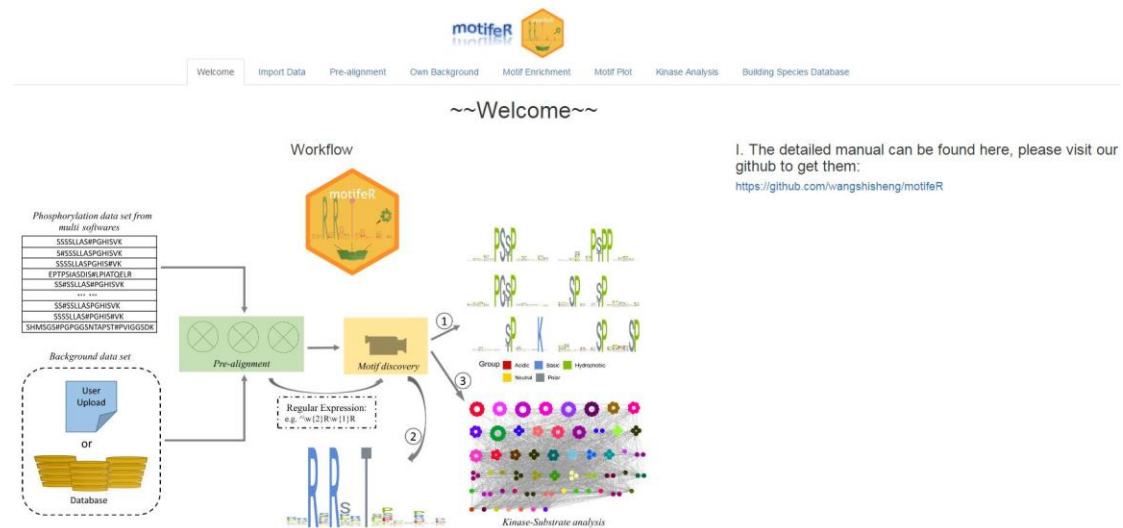


Figure S2. Distribution of peptide modification site number. This plot can be obtained in “Pre-alignment” module.

Supplementary notes:

motifeR a powerful and comprehensive web server, can provide three main functions (sequence pre-alignment, motif discovery, kinase-substrate analysis) to help users find the patterns of residues along the short span of a protein or polypeptide and process kinase-substrate analysis subsequently. Here we present the detailed introduction and operation of motifeR, users can follow this manuscript to analyze their own data freely and conveniently.

Users can visit this site: <https://www.omicsolution.org/wukong/motifeR>. Then the website homepage can be shown like this:



1. Data Preparation

1.1 Foreground data set, which can be obtained from many softwares, such as MaxQuant, Proteome Discovery, Spectronaut and so on. Then users just prepare the peptide sequences like this:

	A
1	Pep
2	TQVPASVESQK#PR
3	LIDGK#DK#AAR
4	VSSPPPTITQQCK#K
5	VAHPQDPHHSSEK#PVIHCHK
6	DLTDYLMK#ILTR
7	DIK#EKLCYVALDFENEMATAASSSSLEK
8	EK#LCYVALDFENEMATAASSSSLEK
9	EITALAPSTMK#IK
10	EITALAPSTMKIK#
11	DSYVGDEAQSK#R
12	RGILTLK#YPIEHGIITNWDDMEK
13	LNKDDPIGNINLAMEIAEK#HLDIPK
14	TPEK#TMQAMQK
15	QSILAIQNEVEK#VIQSYSIR
16	ELPPDQAQYCIK#R
17	SEFYK#HIVLSGGSTMYPGLPSRLER
18	SALSCHLETVILGLLK#TPAQYDASELK
19	SVCHLQK#VFER
20	QTFK#SHFGR
21	LILGLMMPFAHYDAKQLK#
22	GACTDEK#TLTR
23	GSYNIK#SR
24	LHTEGDK#AFVEFLTDEIKEEK

The length of these sequences can be different, and the standard sequences can be obtained in the pre-alignment step. On the other hand, users should mark those modified residues with some label they like (such as “#” or “@”).

1.2 Database/Background data set. motiferR supports users to upload their own database, if they don't want to use the default data. Herein the database formats can be .fasta, .xlsx, .xls, .csv, or .txt. Particularly, if users want to upload the .fasta file, it should contain all protein sequences of the species they study, otherwise, the other formats (.xlsx, .xls, .csv, .txt.) should contain standard sequences like below:

Windows
____MSKVFKKTS
SKVFKKTSSNGKLSI
KVFKKTSSNGKLSIY
TSSNGKLSIYLGKRD
LQVVPAAESSSPQGGL
QVVPAAESSSPQGGLT
VVPAESSSPQGGLTV
MVTNLPCSVTLQPGP
GIDFEVKSFCAENPE
ENPEETVSKRDYVRL
PEAGPGPSAQTIIRRF

2. Import data.

2.1 Uploading data. When users prepare their data (Foreground and Background data set), they can upload these data from here:

The screenshot displays the 'Import Sequence Data' interface of the MotifTool web application. The top navigation bar includes links for 'Welcome', 'Import Data', 'Pre-alignment', 'Own Background', 'Motif Enrichment', 'Motif Plot', 'Kinase Analysis', and 'Building Species Database'. The 'Import Data' tab is active.

The interface is divided into two main sections: 'Import Sequence Data' on the left and a data preview table on the right.

Import Sequence Data Section:

- 1. Upload:** Includes a 'File format' dropdown (set to 'xlsx'), a 'Browse...' button, and a 'No file selected' message.
- 2. Paste:** Includes a 'Paste your data here:' text area containing a list of sequences: 'INSWAPSSWPIK', 'MLISAVSPEIR', 'KINSWAPSSWPIK', 'KINSWAPSSWPIK', and 'INSWAPSSWPIK'.
- Header:** A checkbox labeled 'Header?' is checked.
- First column:** A checkbox labeled 'First column?' is unchecked.
- Sheet index:** A text input field containing the value '1'.
- Central amino acid:** A text input field containing the value 'S'.

Data Preview Table:

The table shows 10 entries. The first column is labeled 'Peptide' and the second is 'Input_ID'. The sequences are: 1. INSWAPSSWPIK, 2. MLISAVSPEIR, 3. KINSWAPSSWPIK, 4. KINSWAPSSWPIK, 5. INSWAPSSWPIK, 6. EGSQGEPTWTPTANLK, 7. EGSQGEPTWTPTANLK, 8. SHMSGSPGPGGNTAPSTPVGGSDKPGMEEK, 9. SHMSGSPGPGGNTAPSTPVGGSDKPGMEEK, 10. SSWSKLLASPGHISVK. The table is paginated, showing 'Showing 1 to 10 of 3,185 entries'.

There are two choices for users:

1. Upload, users choose the right format and then click “Browse” button to import there data; *Header*: this means whether the first row is column names. If true, you should choose this parameter.

First column: this means whether the first column is row names. If true, you should choose this parameter.

2. Paste, users can also paste their sequences in the box.

2.2 Parameters. There are some basic parameters that users can change based on their own data, shown as below:

The screenshot shows a web form with the following fields and options:

- Central amino acid :** A text input field containing the letter 'S'.
- Label of modification :** A text input field containing the hash symbol '#'. Below this field are two radio buttons: '1. Select' (unselected) and '2. Upload' (selected).
- Width :** A text input field containing the number '7'.
- Minimum number :** A text input field containing the number '20'.
- P-value threshold :** A text input field containing the value '0.000001'.
- Please upload your fasta file :** A section containing a 'Browse...' button and a status indicator that says 'No file selected'.

Central amino acid: the central residue that users want to analyze, for example, phosphorylation motif analysis, can center on phosphorylated S, T or Y residues. If they want to analyze multi motif sites, here should be “STY”.

Label of modification: the label represents modification, users can use some label they like, such as “#”, “@”, in which “#” is recommended. Here is an example:

The diagram illustrates the simplification of a motif. On the left, the text 'EGSQGEPWT[Phospho (STY)]PTANLK' is shown. A red arrow points to the right, where the simplified motif 'EGSQGEPWT#PTANLK' is displayed. The modification '[Phospho (STY)]' is replaced by the hash symbol '#', which is the recommended label.

Width: it is the number of left/right side characters of the central residue.

Minimum number: this threshold refers to the minimum number of times you wish each of your extracted motifs to occur in the data set.

P-value threshold: the p-value threshold for the binomial probability. This is used for the selection of significant residue/position pairs in the motif.

Select or Upload fasta file as background data set: if users want to use the default database, they just select relative species, please note, the default database only have Human with 15 length sequences now, more species data will be implemented in the future. Optionally, users can also upload their own fasta file, no species limits here, but the calculation time would be longer.

3. Pre-alignment

This step means align those peptide sequences with the background database (protein sequences) and force the modified sites/residues to be central sites, then users can get the standard peptide window sequences.

The screenshot shows the MotifR Pre-alignment interface. On the left, there are two checkboxes: "1. Pre-aligned ?" (checked) and "2. Check if containing some regular sequence ?" (checked). Below these is a text input field for a "Regular expression :" containing the pattern `^w{2}Rw{1}R`. A "Calculate" button is at the bottom of this section. On the right, there is a table with 10 rows of peptide sequences and their alignment results. The table has columns: "Pep.upload", "Pep.no", "Pep.index", "Seq.window", "PRO.from.Database", and "PRO.index.from.Data". The table shows 10 entries, with the first entry being "INSAPSSPIK" and the last entry being "SSSSLLASPGHSYK".

Pep.upload	Pep.no	Pep.index	Seq.window	PRO.from.Database	PRO.index.from.Data
1 INSAPSSPIK	INSAPSSPIK	3,7	SCRRKINSAPSSPIKINSAPSSPIKTNKA	ADAVK6	413,417
2 MLISAVSPEIR	MLISAVSPEIR	7	KMLISAVSPEIRNRD	ADAVK6	71
3 KINSAPSSPIK	KINSAPSSPIK	4,8	SCRRKINSAPSSPIKINSAPSSPIKTNKA	ADAVK6	413,417
4 KINSAPSSPIK	KINSAPSSPIK	4,7	SCRRKINSAPSSPIKINSAPSSPIKTNKA	ADAVK6	413,416
5 INSAPSSPIK	INSAPSSPIK	7	KINSAPSSPIKTNKA	ADAVK6	417
6 EGSQGEPTWTANLK	EGSQGEPTWTANLK	9	GSQGEPTWTANLKM	ADAVK6	58
7 EGSQGEPTWTANLK	EGSQGEPTWTANLK	11	QGEPTWTANLKM	ADAVK6	60
8 SHMSGSPGGSTAPSTPVGGSDKPGMEEK	SHMSGSPGGSTAPSTPVGGSDKPGMEEK	6	KSHMSGSPGGSTAPSTPVGGSDKPGMEEK	ADFG88	693
9 SHMSGSPGGSTAPSTPVGGSDKPGMEEK	SHMSGSPGGSTAPSTPVGGSDKPGMEEK	14	PGGGSTAPSTPV	ADFG88	701
10 SSSSLLASPGHSYK	SSSSLLASPGHSYK	2,4	LHDLGRSSSSLLASPDGRSSSLLASPGH	ADFG88	737,739

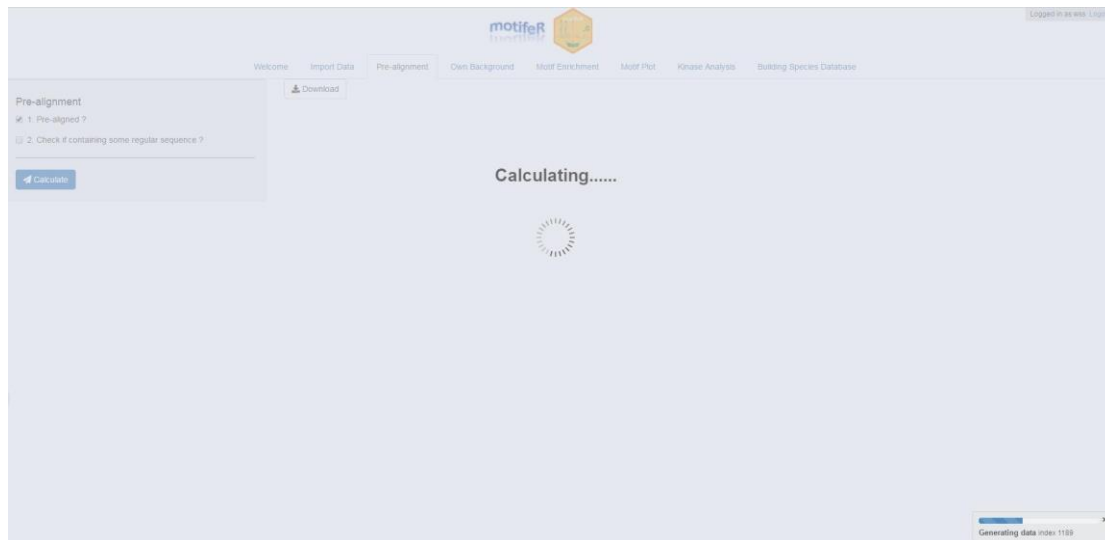
3.1 Parameters

The screenshot shows the "Pre-alignment" form. It has two checkboxes: "1. Pre-aligned ?" (checked) and "2. Check if containing some regular sequence ?" (checked). Below these is a text input field for a "Regular expression :" containing the pattern `^w{2}Rw{1}R`. A "Calculate" button is at the bottom of the form.

Pre-aligned: ask users whether pre-align their sequences, if your sequences are standard (e.g. 15 length amino acids), you can unselect this parameter. Default is true.

Check if containing some regular sequence: if users want to check whether the aligned peptides contain some specific sequences, for example, you want to find those peptides whose 3th and 5th position are R (arginine), then you can select this parameter and type in a simple regular expression, like ' `^w{2}Rw{1}R` ' (more details can be found here: https://en.wikipedia.org/wiki/Regular_expression). Otherwise, you just unselect it.

Then, you can click the "Calculate" button, it should be like this:



A process bar will appear in the bottom right corner to tell users where it goes.

3.2 results

There are two results here.

3.2.1 Alignment results

Pep.upload	Pep.no	Pep.main.index	Pep.all.index	Seqwindows	PRO.from.Database	PROindex.from.Database	Contain.if
1 INS#APSS#PIK	INSAPSSPIK	3.7	3.7	SDRRKINSAPSSPIK KINSAPSSPIKTNKA	ADA/K6	413,417	No
2 MLISA/S#PEIR	MLISA/SPEIR	7	7	KHLISA/SPEIRNRD	ADA/K6	71	No
3 KINS#APSS#PIK	KINSAPSSPIK	4.8	4.8	SDRRKINSAPSSPIK KINSAPSSPIKTNKA	ADA/K6	413,417	No
4 KINS#APSS#PIK	KINSAPSSPIK	4.7	4.7	SDRRKINSAPSSPIK KINSAPSSPIKTNKA	ADA/K6	413,416	No
5 INSAPSS#PIK	INSAPSSPIK	7	7	KINSAPSSPIKTNKA	ADA/K6	417	No
6 EGSGGEPWTP#TANLK	EGSGGEPWTP#TANLK	9	9	GSQGEPTWTP#TANLK	ADA/K6	58	No
7 EGSGGEPWTP#TANLK	EGSGGEPWTP#TANLK	11	11	QGEPTWTP#TANLK	ADA/K6	60	No
8 SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEEK	SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEEK	6	6	IKSHMSGSPGPGGSN	ADFGR8	693	No
9 SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEEK	SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEEK	14	14	PGPGGSNTAPSTPV	ADFGR8	701	No
10 SSSSLASPGH#SVK	SSSSLASPGH#SVK	2.4	2.4	LHDLGRSSSSSLASPDGRSSSSSLASPOH	ADFGR8	737,739	No

Showing 1 to 10 of 3,185 entries

Previous 1 2 3 4 5 ... 319 Next

Pep.upload: this column contains those peptides users upload.

Pep.no: the peptide skeleton.

Pep.index: the position of modified amino acid in the peptide.

Seqwindows: the standard peptides.

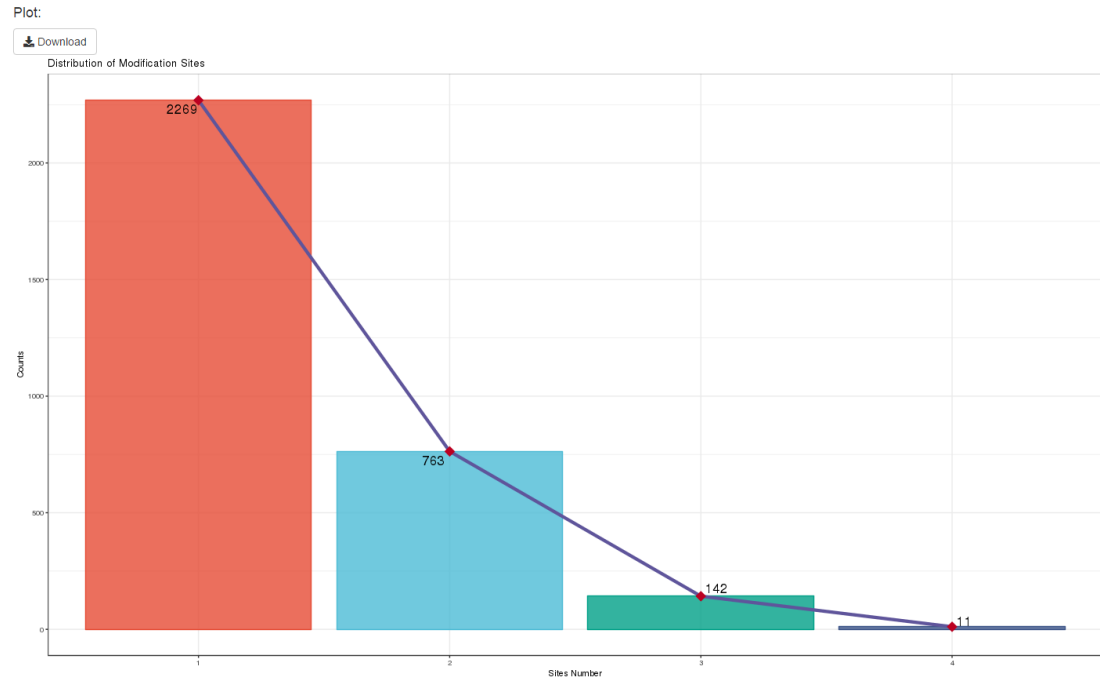
PRO.from.Database: which protein contains this peptide.

PROindex.from.Database: the position of modified amino acid in the protein.

Contain.if: whether containing the sequences that match the regular expression, if true, marked with "Yes", otherwise, "No". This column only appears when users choose the parameter--- *Check if containing some regular sequence*.

3.2.2 Sites number distribution plot

First, this software counts the number of modified sites and plot it:



Second, this tool also extracts those peptides with multi modified sites, the result is similar to that from pre-alignment except the last column.

Multi-Sites Data:

Download

Show 10 entries

Search:

Pep.upload	Pep.no	Pep.index	Seqwindows	PRO.from.Database	PROindex.from.Database	Seqwindows_MultiSites
1	INSAPSSPIK	3,7	SDRRKINSAPSSPIK, KINSAPSSPIKTKNA	ADA/K6	413,417	SDRRKINSAPSSPIK, KINSAPSSPIKTKNA
3	KINSAPSSPIK	4,8	SDRRKINSAPSSPIK, KINSAPSSPIKTKNA	ADA/K6	413,417	SDRRKINSAPSSPIK, KINSAPSSPIKTKNA
4	KINSAPSSPIK	4,7	SDRRKINSAPSSPIK, KINSAPSSPIKTKN	ADA/K6	413,416	SDRRKINSAPSSPIK, KINSAPSSPIKTKN
10	SSSSLLASPGHISVK	2,4	LHDLGRSSSSLLASPG, DLGRSSSSLLASPGH	ADFGR8	737,739	LHDLGRSSSSLLASPG, DLGRSSSSLLASPGH
11	SHMSGSPGPGGSNTAPSTPVI, GSGSDKPKMEEK	14,18	PGPGGSNTAPSTPVI, GSGSDKPKMEEK	ADFGR8	701,705	PGPGGSNTAPSTPVI, GSGSDKPKMEEK
13	SHMSGSPGPGGSNTAPSTPVI, GSGSDKPKMEEK	4,18	TSIKSHMSGSPGPGG, GSGSDKPKMEEK	ADFGR8	691,705	TSIKSHMSGSPGPGG, GSGSDKPKMEEK
14	SSSSLLASPGHISVK	1,2	GLHDLGRSSSSLLAS, LHDLGRSSSSLLASPG	ADFGR8	736,737	GLHDLGRSSSSLLAS, LHDLGRSSSSLLASPG
15	SSSSLLASPGHISVK	1,3	GLHDLGRSSSSLLAS, LHDLGRSSSSLLASPG	ADFGR8	736,738	GLHDLGRSSSSLLAS, LHDLGRSSSSLLASPG
16	SHMSGSPGPGGSNTAPSTPVI, GSGSDKPKMEEK	4,6	TSIKSHMSGSPGPGG, IKSHMSGSPGPGGSN	ADFGR8	691,693	TSIKSHMSGSPGPGG, IKSHMSGSPGPGGSN
17	SHMSGSPGPGGSNTAPSTPVI, GSGSDKPKMEEK	12,18	GSPGPGGSNTAPSTPVI, GSGSDKPKMEEK	ADFGR8	699,705	GSPGPGGSNTAPSTPVI, GSGSDKPKMEEK

Showing 1 to 10 of 916 entries

Previous 1 2 3 4 5 ... 92 Next

Seqwindows_MultiSites: the modified amino acid will be replaced with “X” if it is not the central residue, for example, ‘YES#DEDS#LGSSGR’, when the 3th amino acid is considered as central residue, the 7th amino acid will be replaced with “X”, thus the standard sequence is ‘YSNRKYESDEDXLGS’, conversely, the standard sequence should be ‘KYEXDEDSLGSSGRV’.

4. Own Background

Users can upload their own background database, but it is noteworthy that the database must contain peptide sequences with standard length, not protein sequences.

motifeR
motifEr

Welcome Import Data Pre-alignment **Own Background** Motif Enrichment Motif Plot

☒ Upload your own background data ?

File format :
☒ .xlsx ☐ .xls ☐ .csv/txt

Import your data :
Browse... No file selected

☐ Header ?
☐ First column ?

Sheet:
1

Users can click “Browse” button and import their data, like this:

Welcome Import Data Pre-alignment **Own Background**

Show 10 entries

☒ Upload your own background data ?

File format :
☐ .xlsx ☐ .xls ☒ .csv/txt

Import your data :
Browse... Seqwindows_database.csv
Upload complete

☒ Header ?
☐ First column ?

Separator :
☒ Comma
☐ Semicolon
☐ Tab
☐ BlankSpace

	Seqwindows
1	SDRRKINSAPSSPIK
2	KINSAPSSPIKTNKA
3	KMLISAVSPEIRNRD
4	SDRRKINSAPSSPIK
5	KINSAPSSPIKTNKA
6	SDRRKINSAPSSPIK
7	RKINSAPSSPIKTNK
8	KINSAPSSPIKTNKA
9	GSQGEPWTP TANLKM
10	QGEPWTP TANLKMLI

Showing 1 to 10 of 75,547 entries

Please note, if you upload you own fasta file as background database, you should unselect this parameter. If you choose this parameter stubbornly, this software will take the data in this step as background database and ignore that you upload before.

5. Motif Enrichment

This step will find overrepresented sequence motifs.

5.1 Parameters

☒ Species data as background ?

☐ Only use multi-site data ?

Calculate

Species data as background: if you upload your own fasta file as background database in the 'Import Data' step, you can ignore this parameter (select or unselect is same). Otherwise, if you choose the database in our system in the 'Import Data' step, selecting this parameter means this software will take the database in our system as background database. If you don't choose, the software will take the foreground data as background database.

Only use multi-site data: if selected, this tool will only take the peptides with multi modification sites as foreground data, that is, it will use the sequences in the *Seqwindows_MultiSites* column obtain from 'Pre-alignment' step as foreground data.

5.2 Results

The enrichment results like this:

Welcome Import Data Pre-alignment Own Background Motif Enrichment Motif Plot Kinase Analysis Building Species Database

Multiple motifs Regular sequence motif

Download

Show 10 entries Search:

motif	score	fg.matches	fg.size	bg.matches	bg.size	fold.increase	Enrich.seq
1P.SP.....	317.89810602509	120	1726	7499	902328	8.36568552902457	KINSAPSSPIKTNKA:NVGSPKSPTHASPO:NSAVTLPSPGSSPFP:VQWLNQSPSTTTSSN:PKK
2SP.....	307.65265568589	451	1606	57796	894629	4.34783714946362	KMLISAVSPEIRNRD:IKSHMSGSPGPGGSN:SSSSLLASPGHISVK:KKKKNRHSDPHGMG:VKC
3SSP.....	615.305311137178	70	1155	7774	837033	6.52550459573871	RKINSAPSSPIKTNK:VTAEADSSSPTGILA:TSSLDLS:SSPSVTT:FSKERSPSSPVVVK:SKETC
4S.SP.....	318.541773931106	47	1085	6581	829259	5.45841337686974	TSIKSHMSGSPGPGG:LKSPVSESVSPVVPD:FPEPTCLASAPPNAP:PGTPYKVCSCPTSGA:SG
5R.S.....	307.65265568589	160	1038	47281	822678	2.68204375928299	HDLGRSSSSLLASPG:PPQARTSSLDNEGPH:TAGCRGSSAVLNVT:VPLRRRHSEQVANGP:W
6SD.D.....	24.6017271480482	29	878	2832	775397	9.04345432287042	PVPPETPSDSOHKKK:SAGYEEISDPDMEEK:LKRRLSYSDSLKRA:EDGEEDSDSDYEIS:TLQ
7SD.E.....	22.3583226679643	30	849	4154	772565	6.57176615497685	EGEEDSDSDYEISAK:GEAPEPDSDAEVAEA:SRFFTTGSDSESESS:FHYRTLHSDDEGTVL:GE
8S.S.....	10.4709411486104	150	819	81386	768411	1.72922388844046	DLGRSSSSLLASPGH:PPETPSDSOHKKKKK:GLVYKSGSGEIGSET:LATSEKSMFVLGSV:PKL
9S.S.....	10.4914122398175	130	669	73679	687025	1.81194934573683	EPTPSIASDISLP:IA:GRKTSIKSHMSGSPG:PSIASDISLP:ATQE:NTLKSVPVSESV:VVT:TPTSSLC
10S...D...	8.09910955740197	58	539	28674	613346	2.30174116480277	DLEVFRNSLYAPDYS:KNSMPTVSFLDQDQS:GKQPLLLSEDEEDTK:RLHGGFDSDCSEDEGE:GI

Showing 1 to 10 of 13 entries

Previous 1 2 Next

motif: the overrepresented motif.

score: the motif score, which is calculated by taking the sum of the negative log probabilities used to fix each position of the motif. Higher motif scores typically correspond to motifs that are more statistically significant as well as more specific.

fg.matches: frequency of sequences matching this motif in the foreground set

fg.size: total number of foreground sequences

bg.matches: frequency of sequences matching this motif in the background set

bg.size: total number of background sequences

fold.increase: An indicator of the enrichment level of the extracted motifs. Specifically, it is calculated as $(\text{foreground matches}/\text{foreground size})/(\text{background matches}/\text{background size})$.

Enrich.seq: those peptides are overrepresented in this motif.

Enrich.pro: those proteins in which the peptides exist from *Enrich.seq*.

6. Motif Plot

This step will plot the motifs from the enrichment results.

If users only input one number in the 'Motif index for plot' parameter, it will plot the relative motif, shown as below:



If users type in '1-10', it will plot the 1th to 10th motifs, like this:



7. Kinase-Substrate Analysis

This step will take a formatted phosphoproteomics data input and perform kinase-substrate analysis calculations to infer relative kinase activities. Users should note here:

- There are only Human database for this analysis in this system at present.
- This is only for phosphoproteomics data, other modification data are not inappropriate here.

	KIN_ACC_ID	SUB_ACC_ID	Motif	KINASE	SUBSTRATE	networkin_score
1	O43683	Q12834	...PSP...	BUB1	CDC20	300
2	O43683	O43683	...PSP...	BUB1	BUB1	300
20	O43683	P0C058	...PSP...	BUB1	H2A 1	300
21	O43683	O43684	...PSP...	BUB1	BUB3	300
28	O43683	Q12834	...SP...	BUB1	CDC20	300
29	O43683	O43683	...SP...	BUB1	BUB1	300
47	O43683	P0C058	...SP...	BUB1	H2A 1	300
48	O43683	O43684	...SP...	BUB1	BUB3	300
65	O60285	O14974	...SSP...	Nuak1	MYPT1	300
66	O60285	O55835	...SSP...	Nuak1	LATS1	300

7.1 Parameters

minimum NetworkKIN score:

Select Motif class:

Calculate

minimum NetworkKIN score: a numeric value between 1 and infinity setting the minimum NetworkKIN score.

Select Motif class: select those motifs you want to analyze, these motifs are obtained from 'Motif Enrichment' step:

Select Motif class:

- ...SSP...
- ...S...E...
- ...S.S...
- ...R.S...
- ...S.SP...
- ...SD.E...
- ...S...P.
- S.S

7.2 Results

First, this software give a result table containing kinases and substrates information.

Welcome Import Data Pre-alignment Own Background Motif Enrichment Motif Plot Kinase Analysis Building Species Database

Results Network Plot

Download

Show 10 entries Search:

	KIN_ACC_ID	SUB_ACC_ID	Motif	KINASE	SUBSTRATE	networkin_score
1	O43683	Q12834PSP....	BUB1	CDC20	300
2	O43683	O43683PSP....	BUB1	BUB1	300
20	O43683	P0C0S8PSP....	BUB1	H2A.1	300
21	O43683	O43684PSP....	BUB1	BUB3	300
28	O43683	Q12834SP....	BUB1	CDC20	300
29	O43683	O43683SP....	BUB1	BUB1	300
47	O43683	P0C0S8SP....	BUB1	H2A.1	300
48	O43683	O43684SP....	BUB1	BUB3	300
55	O60285	O14974SSP....	NuaK1	MYPT1	300
56	O60285	O95835SSP....	NuaK1	LATS1	300

Showing 1 to 10 of 776 entries Previous 1 2 3 4 5 ... 78 Next

KIN_ACC_ID: kinase uniprot id.

SUB_ACC_ID: substrate uniprot id.

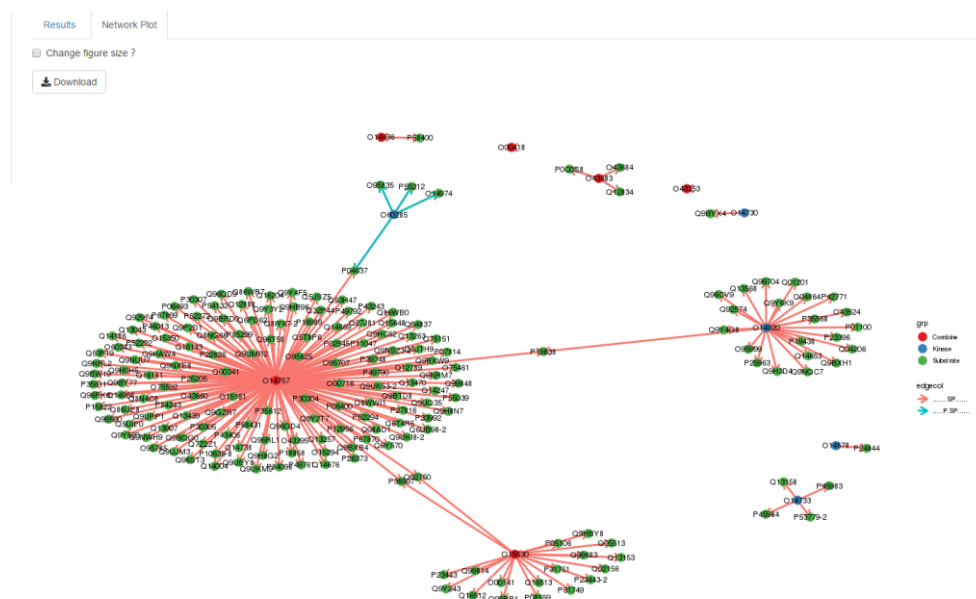
Motif: the overrepresented motif.

KINASE: kinase gene name.

SUBSTRATE: substrate name.

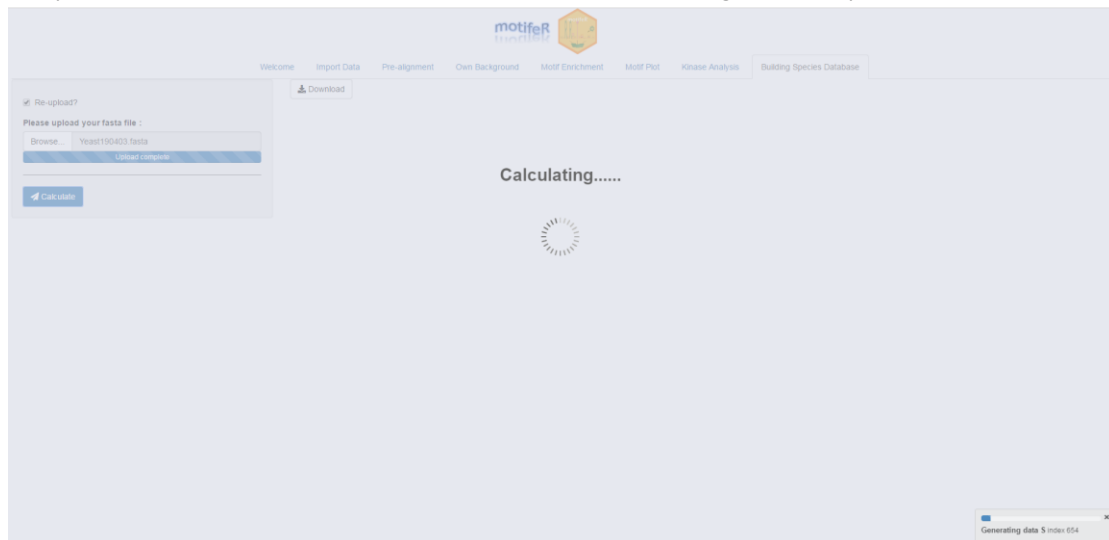
networkin_score: the prediction score from networkIN (<https://networkin.info/>).

Second, this tool will plot the kinase-substrate network:



8. Building Species Database

This step can build the standard database based on the fasta file that users upload, herein there is no species limit. And this results can also be used in 'Own background' step.



If users want to build their own database, they can select the 'Re-upload' parameter and then upload a fasta file, the results will be shown as below:

Welcome

Import Data

Pre-alignment

Own Background

Motif Enrichment

Motif Plot

Kinase Analysis

Building Species Database

Download

Show

10

▼

entries

Search:

	ID	Windows	Center
1	P33755	_MLIRFRSKNGTHRV	S
2	P33755	KNGTHRVSCQENDLF	S
3	P33755	GQGIHVAISELADRTV	S
4	P33755	DMLILNYSDKPFANEK	S
5	P33755	GVNVEIGSVGIDSKG	S
6	P33755	IGSVGIDSKGIQHR	S
7	P33755	GLIPRQKSKLCKHGD	S
8	P33755	RGMCEYCSPLPPWDK	S
9	P33755	KNKIKHISFHSYLKK	S
10	P33755	IKHISFHSYLKKLNE	S

Showing 1 to 10 of 274,123 entries

Previous

1

2

3

4

5

...

27413

Next

ID: uniprot ids.

Windows: the standard peptides.

Center: Center residue.

Those basic parameters in 'Import Data' step are also usable here. Therefore, if you want to get different results, just change those parameters.

All results can be saved to .pdf or .csv files by clicking corresponding "Download" button.