

MotifStack

A TOOL TO VISUALIZE SEQUENCE LOGO ALIGNMENTS

Jianhong Ou
Julie Zhu



INSTALL THE WORKSHOP PKG

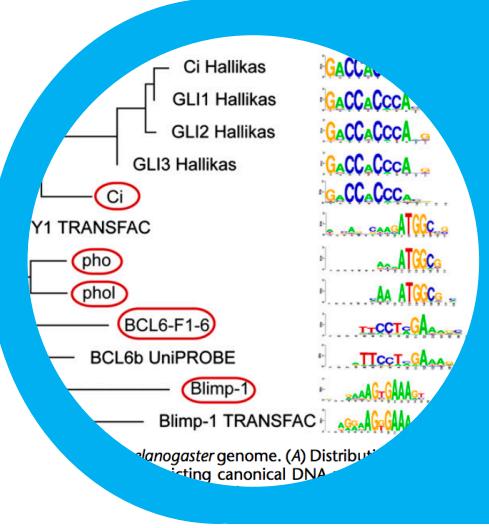
```
## set the working directory,  
## replace "~/Downloads/workshop2020" by your path  
wd <- "~/Downloads/workshop2020"  
dir.create(wd)  
setwd(wd)  
library(BiocManager)  
install("jiahong/workshop2020", build_vignettes = TRUE)  
vignette("motifStack", package="workshop2020")
```

<https://github.com/jiahong/workshop2020>
<https://bioconductor.org/packages/motifStack>

Slides:

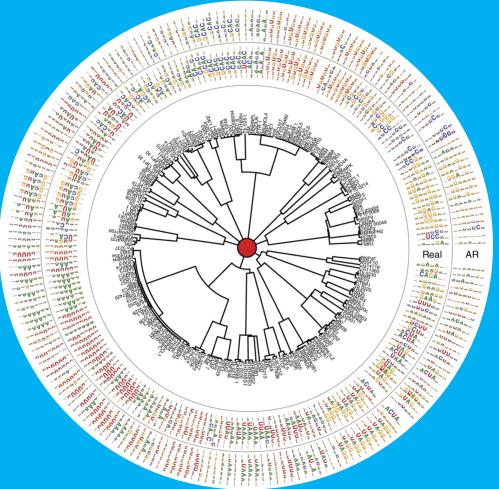
https://github.com/jiahong/workshop2020/blob/master/inst/extdata/motifStack_workshop2020.pdf





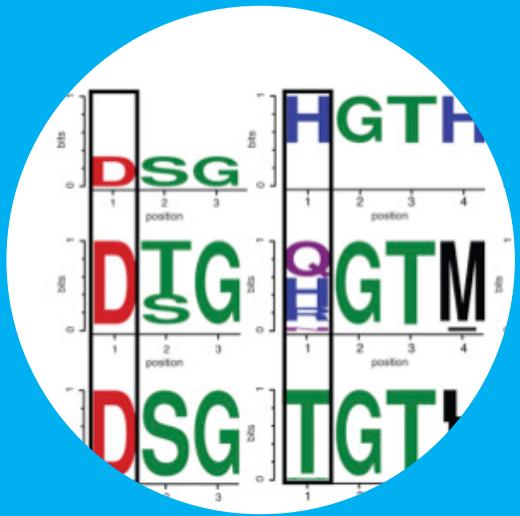
2012 motifStack package initialed

Enuameh MS, et al. 2013. doi: [10.1101/gr.151472.112](https://doi.org/10.1101/gr.151472.112)



2015 compare 130 RBPs for two algorithms

Pelossof R, et al. 2015. doi: 10.1038/nbt.3343



2016 plot Amino Acid (AA) logo

Kevorkian et.al., 2016. doi: [10.1016/j.biochi.2015.07.023](https://doi.org/10.1016/j.biochi.2015.07.023)

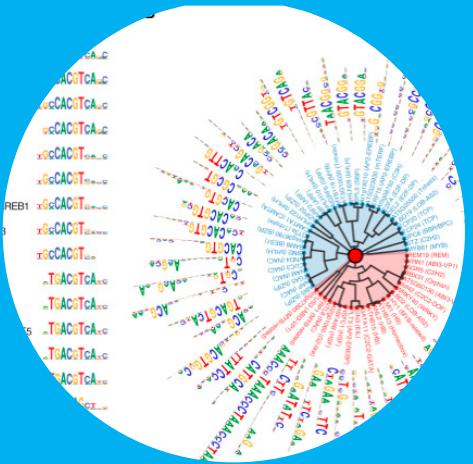
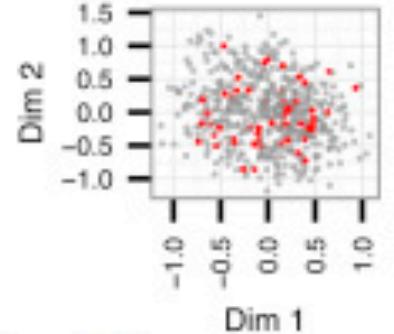
bZIP family

Protein
Phylogeny

- Group S
- Group G
- Group A
- Group D
- Group I

bZIP2_GBF5
bZIP11_GBF6_ATB2
bZIP44
bZIP53
bZIP55_GBF3

-G-CACGTCA-C
-G-CACGTCA-C
-G-CACGTCA-C
-G-CACGTCA-C
-G-GCACGT-

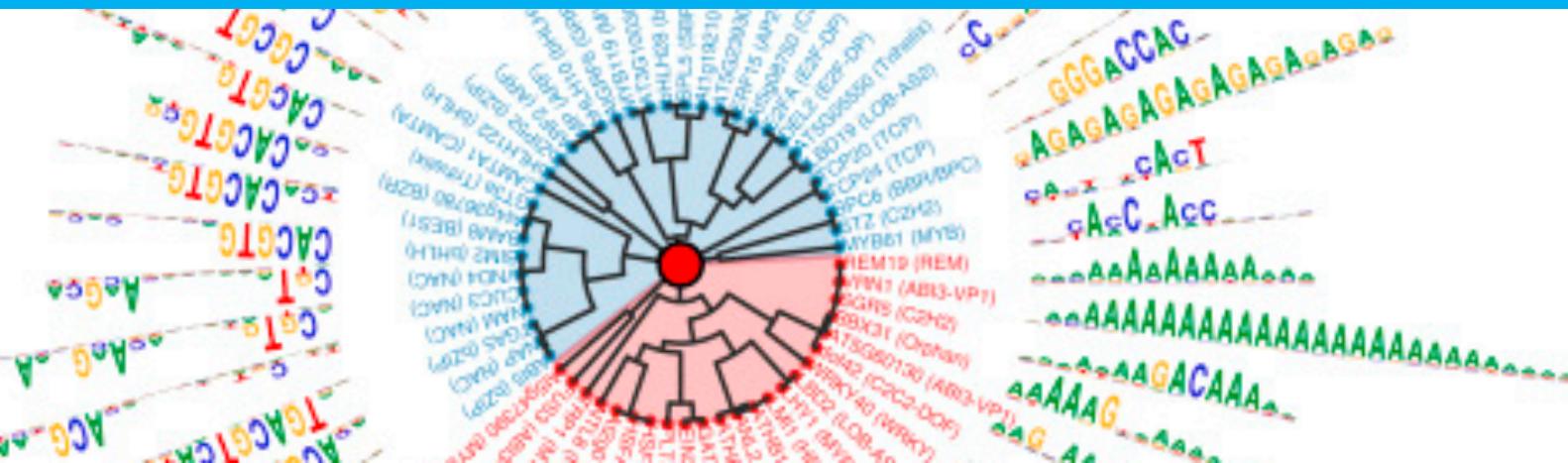


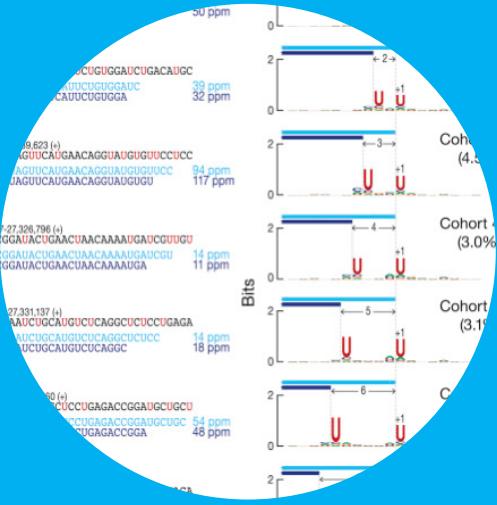
2016 impressed people by its beauty

O'Malley RC, et al. 2016. doi: 10.1016/j.cell.2016.04.038

bZIP39_ABI5
bZIP36_ABF2_AREB1
bZIP66_AREB3
bZIP16
bZIP22_TGA3
bZIP20_TGA2

-G-GCACGT-
-G-GCACGT-
-G-GCACGT-
-G-GCACGT-
-TGACGTCA-
-TGACGTCA-

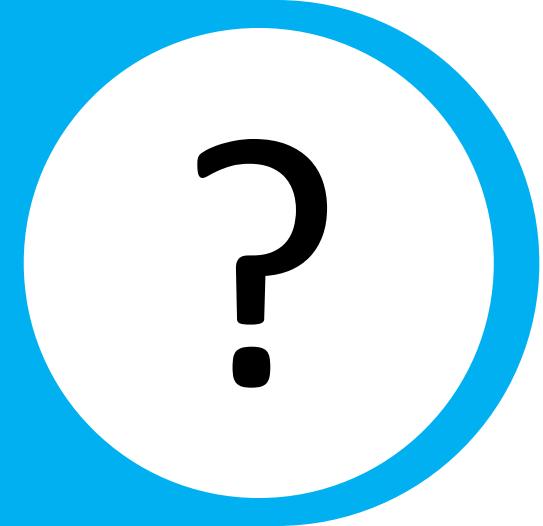




2018 plot piRNAs

Gainetdinov et.al., 2018. doi: [10.1016/j.molcel.2018.08.007](https://doi.org/10.1016/j.molcel.2018.08.007)

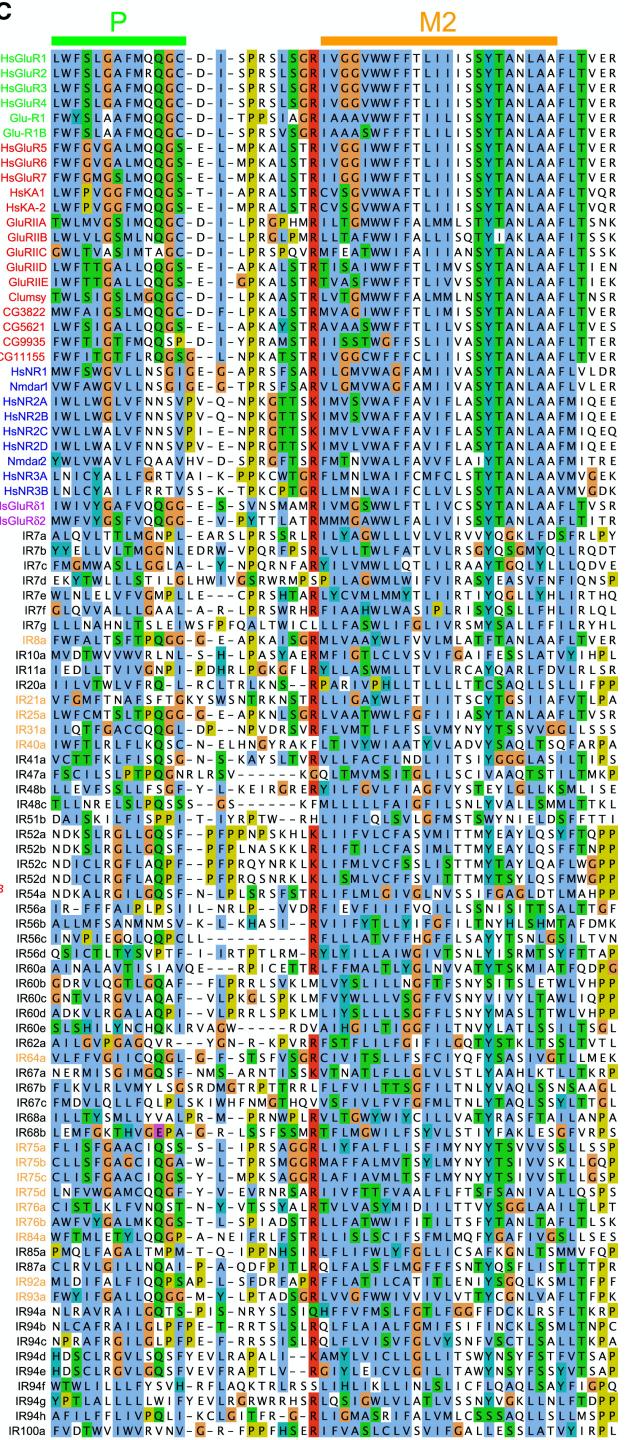
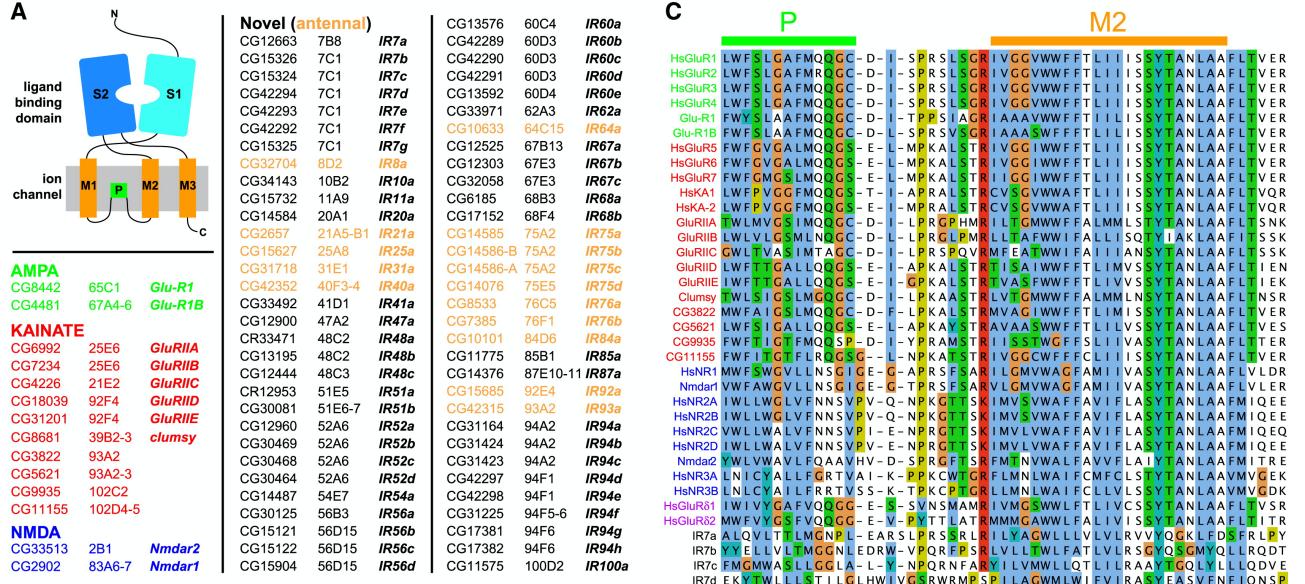
Nishida et.al., 2018. doi: [10.1038/nature25788](https://doi.org/10.1038/nature25788)

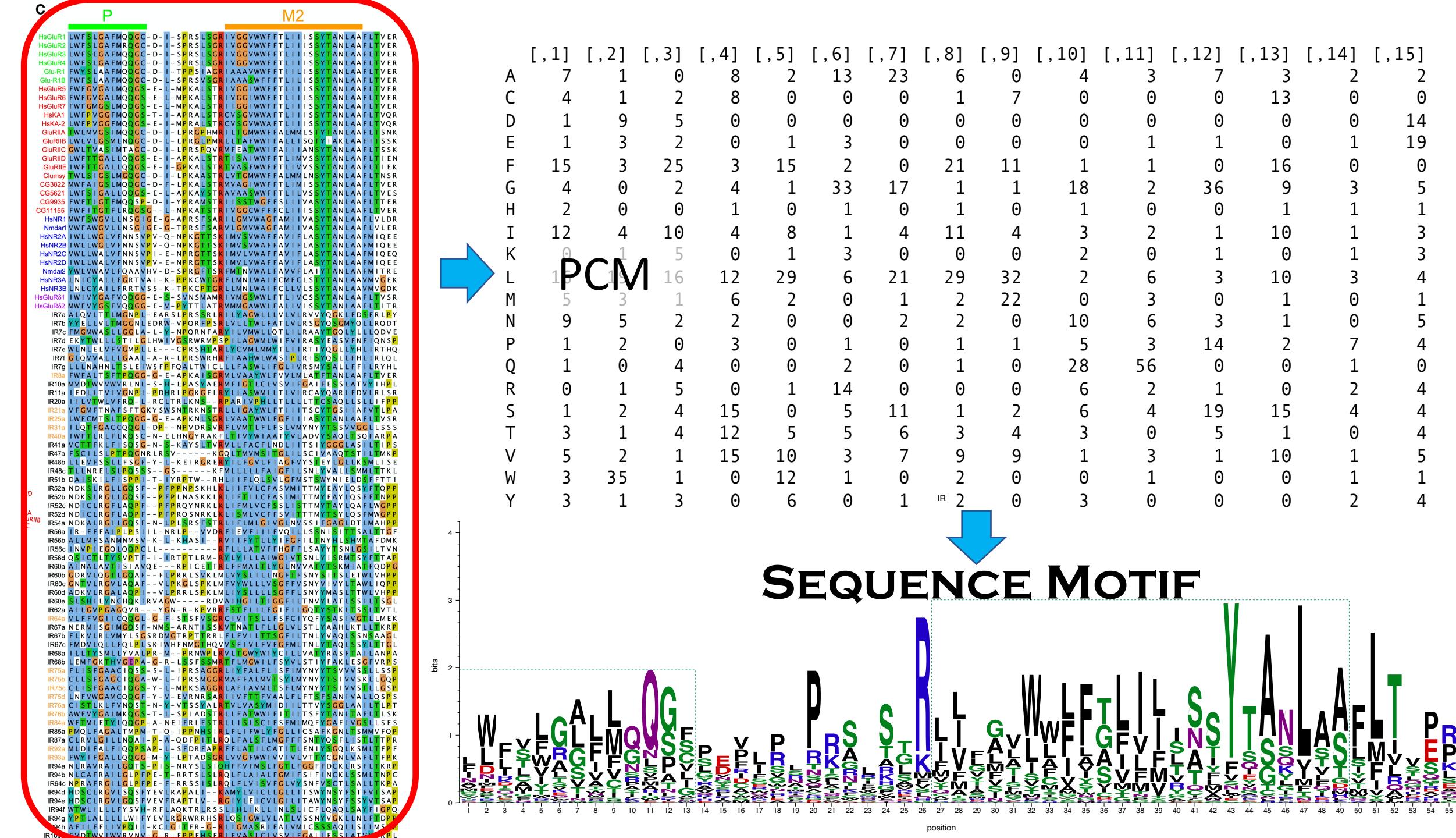


SEQUENCE MOTIF STEREOTYPICAL ELEMENT

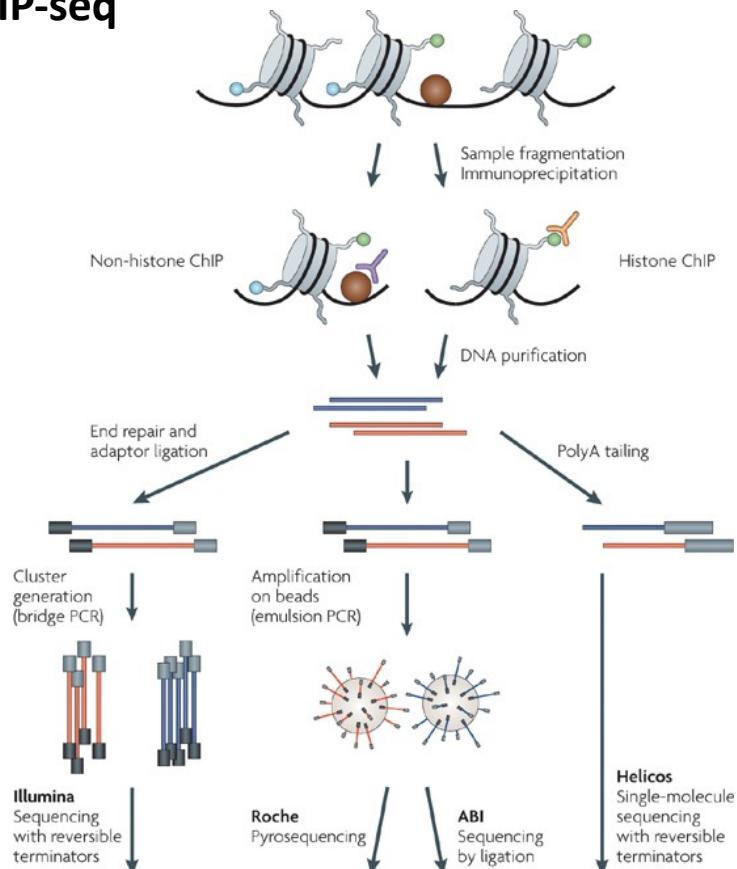


Benton et.al., 2009. doi: 10.1016/j.cell.2008.12.001





ChIP-seq



J²⁰ JASPAR²⁰²⁰

HOCOMOCO

CIS-BP

Park. 2009. doi: 10.1038/nrg2641

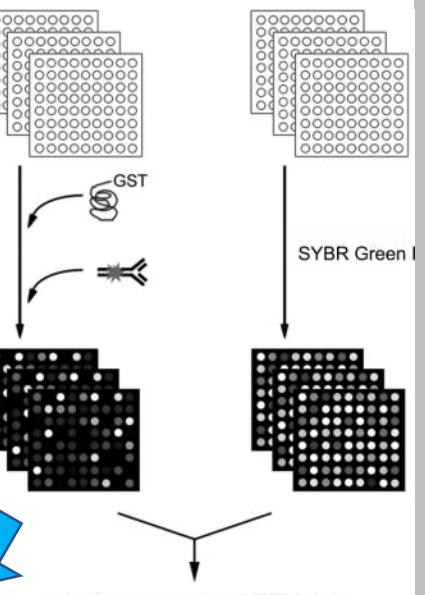
PBM

double-stranded DNA microarrays

bind epitope-tagged TF to dsDNA microarrays

label with fluorophore-tagged anti(epitope) antibody

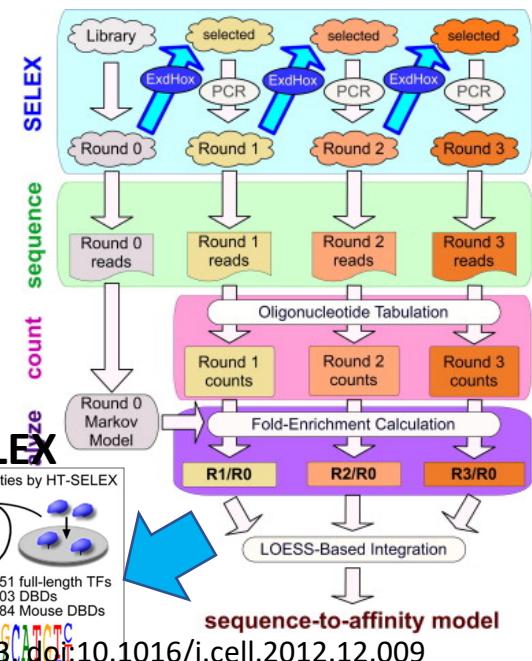
scan triplicate microarrays



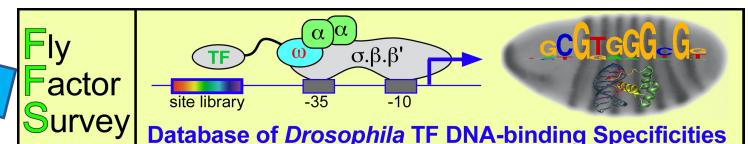
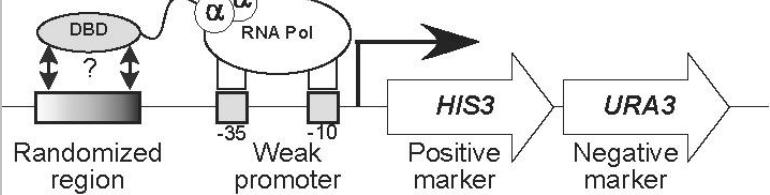
UniPROBE Database

Berger et.al., 2006. doi: 10.1385/1-59745-097-9:245

SELEX

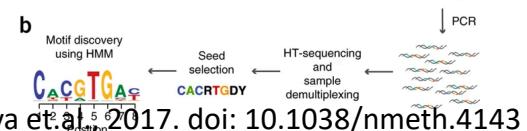
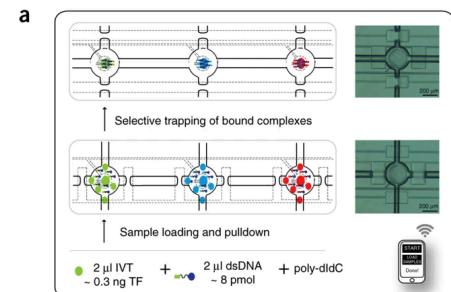


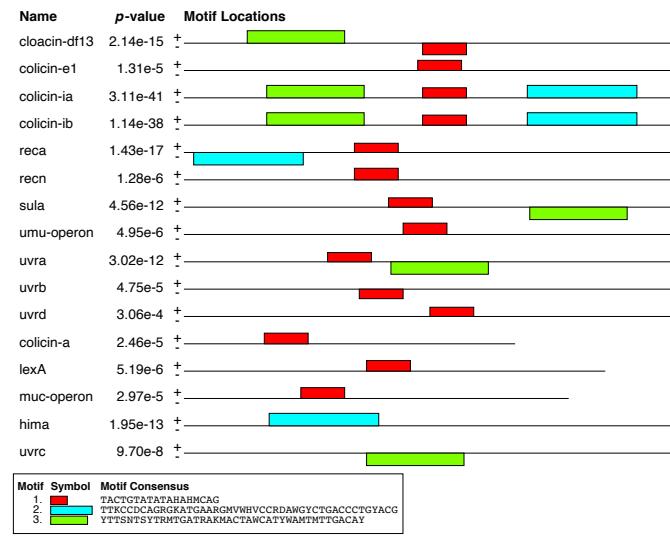
B1H



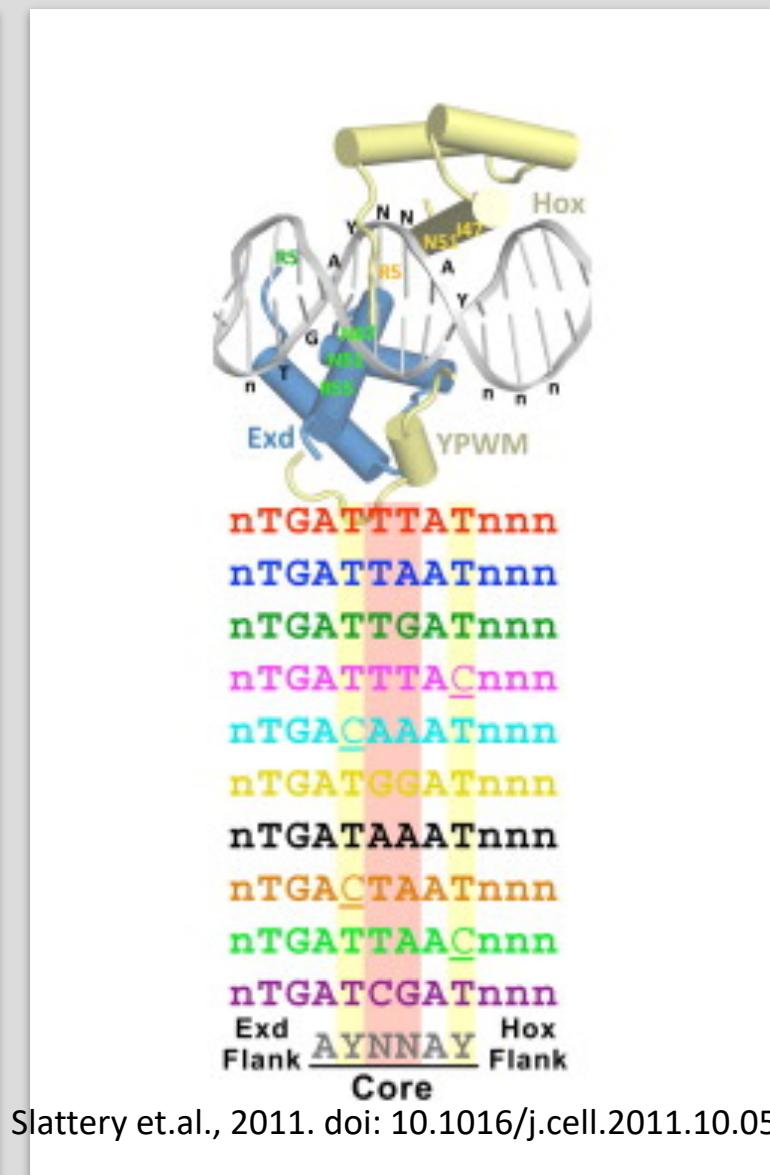
Meng et.al., 2005. doi: 10.1038/nbt1120

SMiLE-seq





Bailey et.al., 2015. doi: 10.1093/nar/gkv416

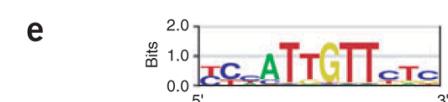


Slattery et.al., 2011. doi: 10.1016/j.cell.2011.10.053

a HEM13 CCCATTGTTCTC
 HEM13 TTTCTGGTTCTC
 HEM13 TCAATTGTTTAG
 ANB1 CTCATTGTTGTC
 ANB1 TCCATTGTTCTC
 ANB1 CCTATTGTTCTC
 ANB1 TCCATTGTTCGT
 ROX1 CCAATTGTTTG

b YCHATTGTTCTC

c A 002700000010
 C 464100000505
 G 000001800112
 T 422087088261



D'haeseleer. 2006. doi: 10.1038/nbt0406-423

PLOT SINGLE SEQUENCE LOGO

WebLogo 3 home create examples manual

Introduction

[**WebLogo**](#) is a web-based application designed to make the generation of sequence logos easy and painless. WebLogo has been featured in over [7000](#) scientific publications.

A [sequence logo](#) is a graphical representation of an amino acid or nucleic acid multiple sequence alignment. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of, for example, a binding site, than would a consensus sequence.

[WebLogo](#) is a web-based application designed to make the generation of sequence logos easy and painless. WebLogo has featured in over 7000 scientific publications

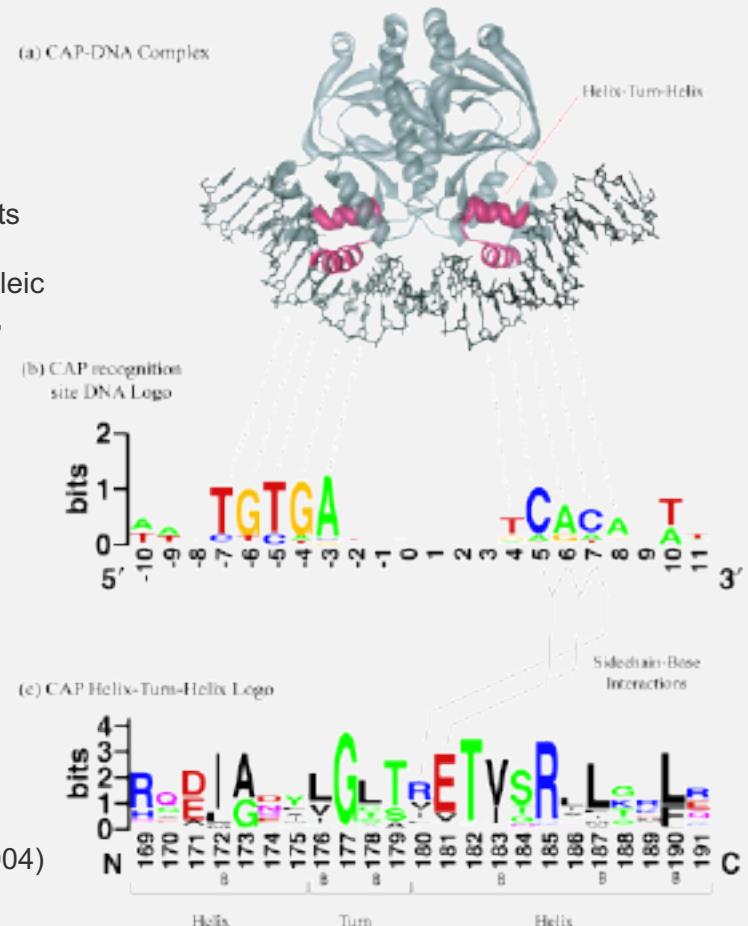
- [Create your own logos](#)
 - [View example sequence logos and input data.](#)
 - [Read the release notes for latest changes and updates](#)
 - [Read the User's Manual](#)
 - [WebLogo source code](#)
 - [WebLogo discussion group](#)

References

Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190, (2004)

[Full Text]

Schneider TD, Stephens RM. 1990. [Sequence Logos: A New Way to Display Consensus Sequences](#). *Nucleic Acids Res.* 18:6097-6100.





Tomtom

Motif Comparison Tool

Version 5.1.1

Data Submission Form

Search one or more motifs against a motif database.

Input query motifs
Enter the motif(s) to compare to known motifs. [?](#)

Type in motifs [▼](#) DNA [▼](#) DNA [?](#)

Select target motifs
Select a motif database or provide motifs to compare with. [?](#)

Eukaryote DNA [▼](#) DNA [?](#)

Vertebrates (In vivo and in silico) [▼](#) [?](#)

Allow alphabet expansion [?](#)

Run immediately
 Search for one motif without queueing [?](#)

Input job details
(Optional) Enter a job description. [?](#)

► Advanced options

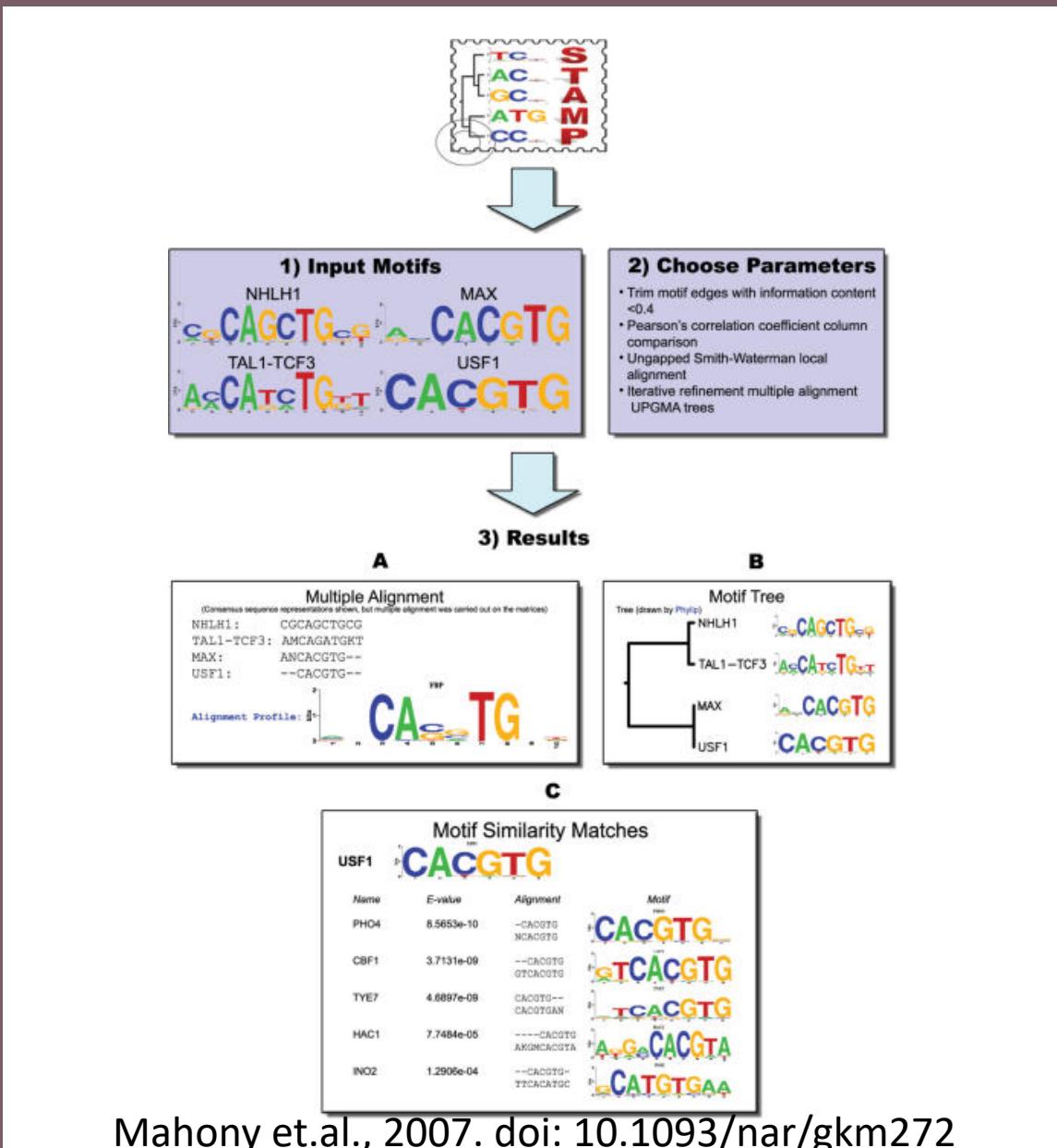
Note: if the combined form inputs exceed 80MB the job will be rejected.

Start Search **Clear Input**

Version 5.1.1 Please send comments and questions to: meme-suite@uw.edu Powered by Opal

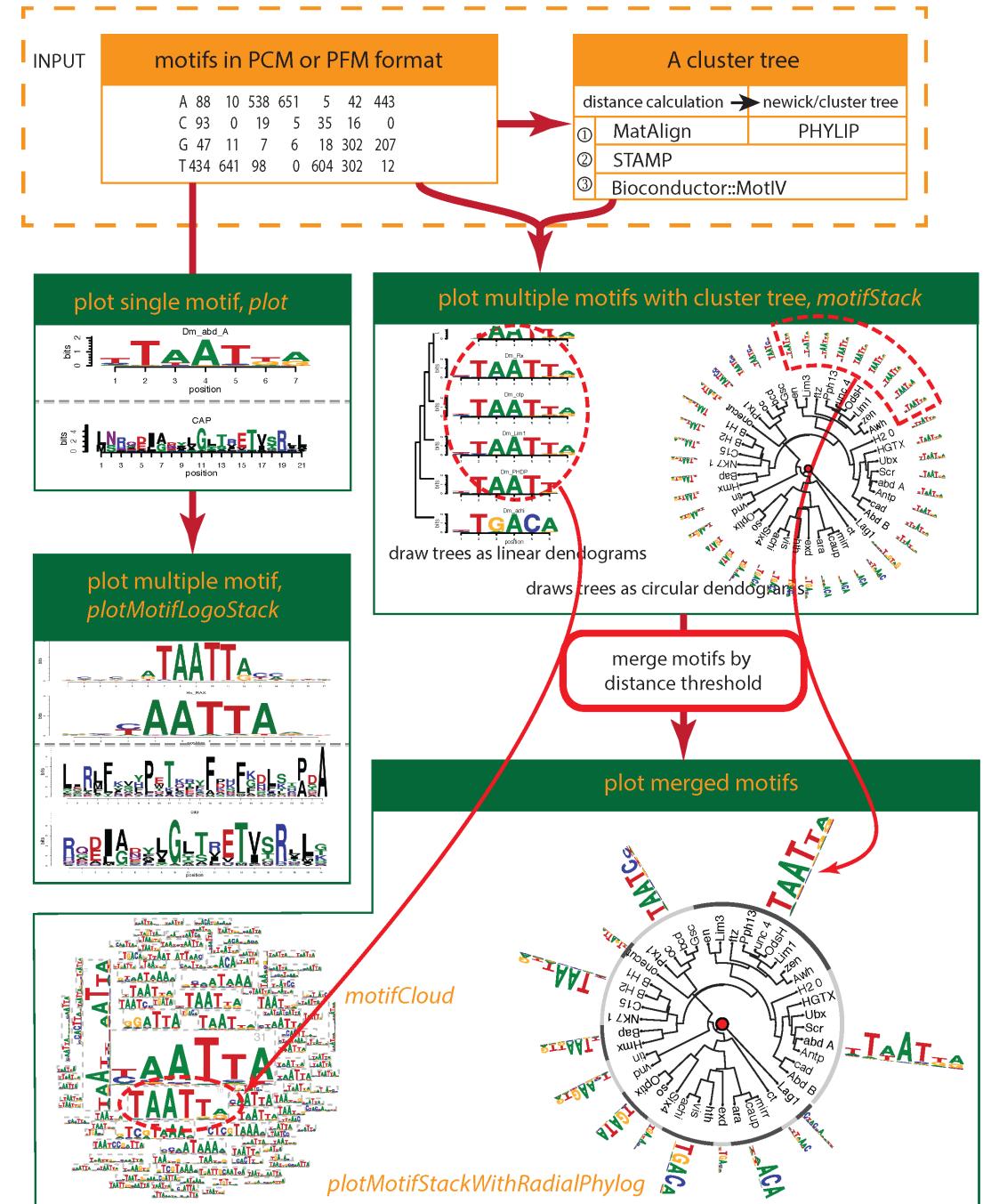
Gupta et.al., 2007. doi: 10.1186/gb-2007-8-2-r24

Tomtom compares one or more motifs against a database of known motifs (e.g., JASPAR). Tomtom will rank the motifs in the database and produce an alignment for each significant match ([sample output](#) for motif and JASPAR CORE 2014 database). See this [Manual](#) for more information.

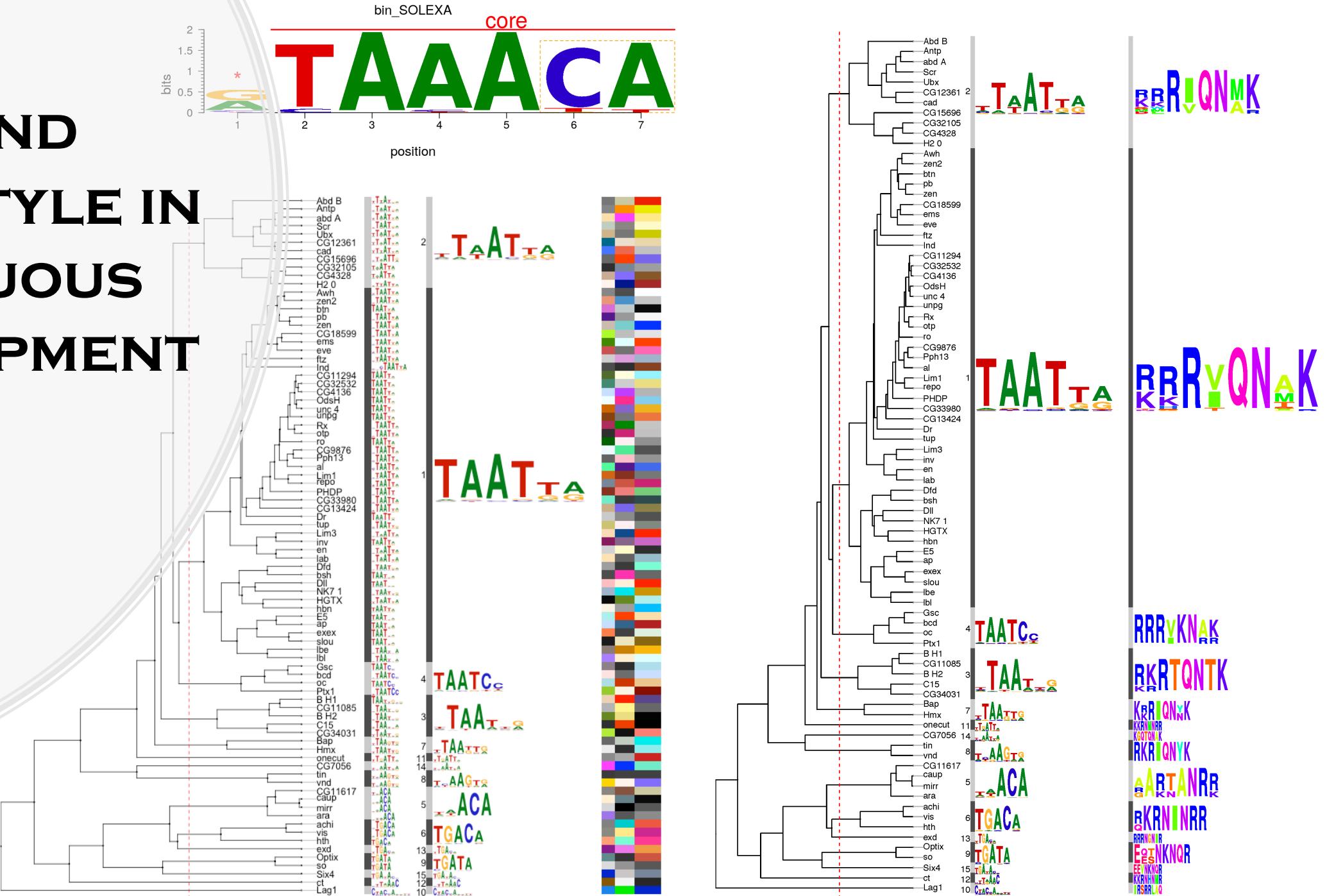


FROM SINGLE MOTIF TO MULTIPLE MOTIFS

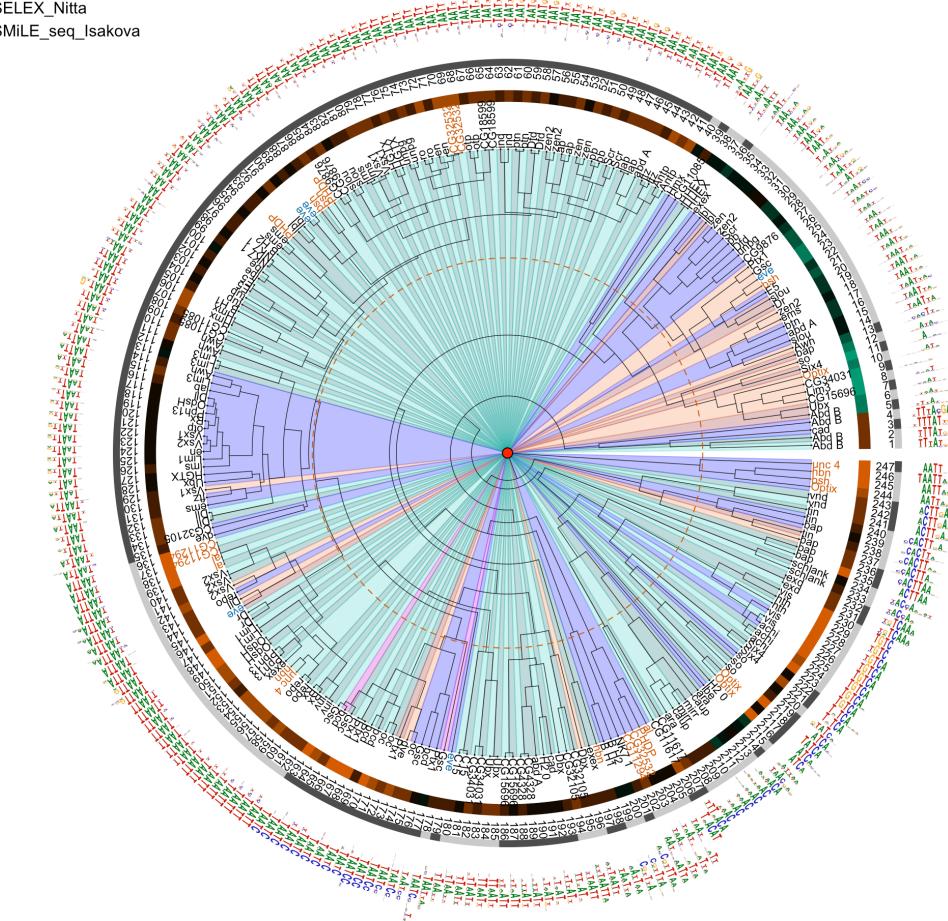
- ✿ Plot aligned motifs
- ✿ Powerful tool to visualize bunch of sequence logos
- ✿ Highlight grouped motifs by their signatures
- ✿ Multiple style and technique to show and label motifs



MORE AND MORE STYLE IN CONTINUOUS DEVELOPMENT



WHAT WILL WE MAKE TODAY



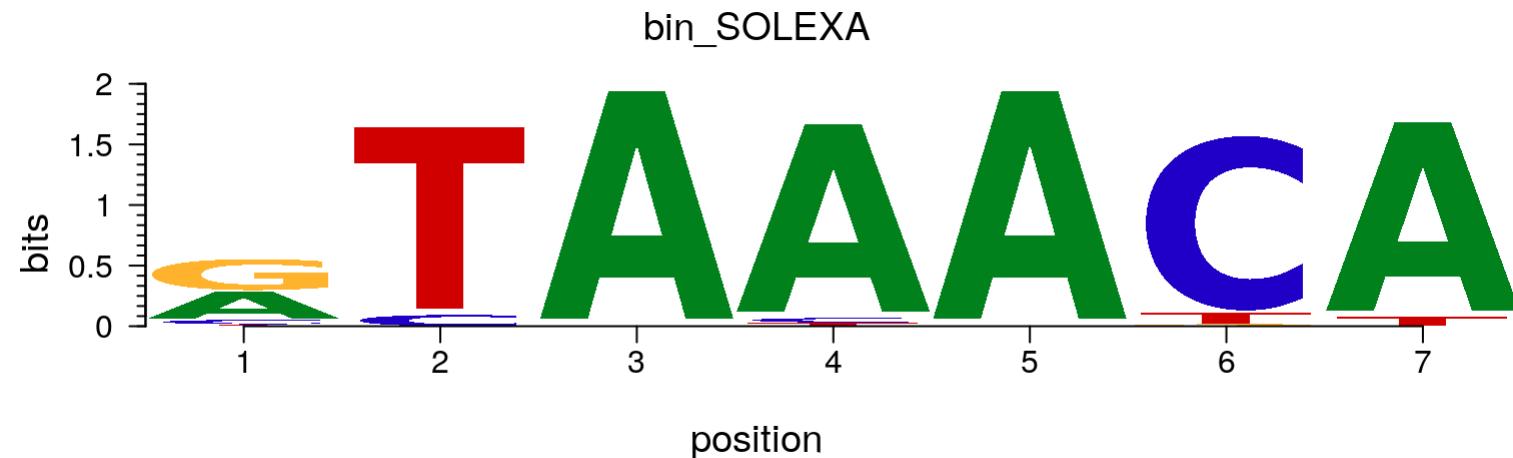
INSTALL motifStack PACKAGE

```
if(packageVersion("motifStack")<"1.33.3"){
  BiocManager::install("jianhong/motifStack", build_vignettes=TRUE)
}
```

- ❖ Starting from version 1.33.2, *motifStack* does not require cario or ghostscript anymore. It will use cario if cario (≥ 1.6) is installed or use ghostscript if gs command is available. Otherwise, *motifStack* will use embed font to plot the sequence logo.
- ❖ MatAlign algorithm was included in *motifStack* since 1.33.2.

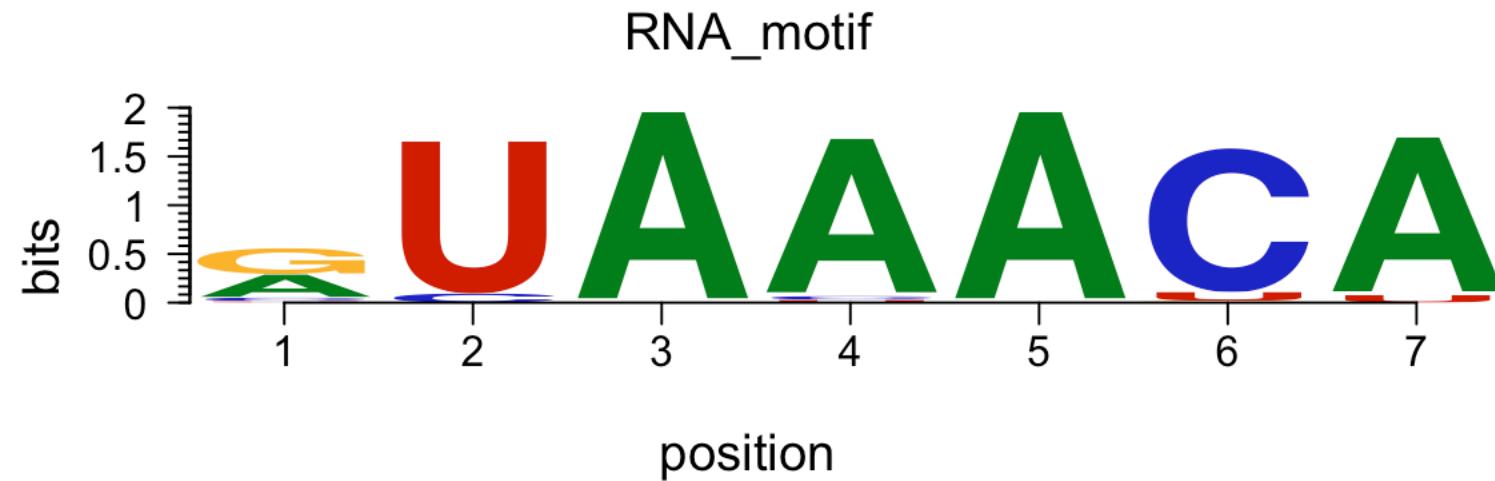
PLOT A DNA SEQUENCE LOGO

```
library(motifStack)
pcm <- importMatrix(system.file(" extdata ", "bin_SOLEXA.pcm" , package = " motifStack "),
                     format = "pcm", to = "pcm")
plot(pcm)
```



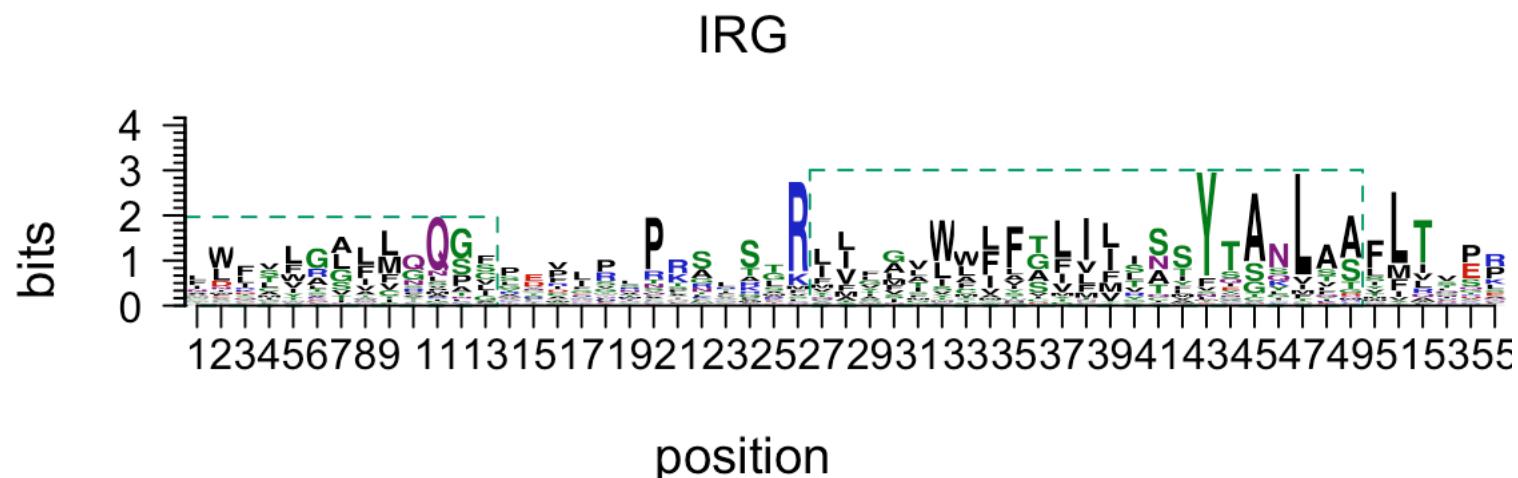
PLOT AN RNA SEQUENCE LOGO

```
library(motifStack)
pcm <- read.table(file.path(find.package("motifStack"), "extdata", "bin_SOLEXA.pcm"))
pcm <- pcm[,3:ncol(pcm)]
rownames(pcm) <- c("A", "C", "G", "U")
motif <- new("pcm", mat=as.matrix(pcm), name="bin_SOLEXA")
plot(motif)
```



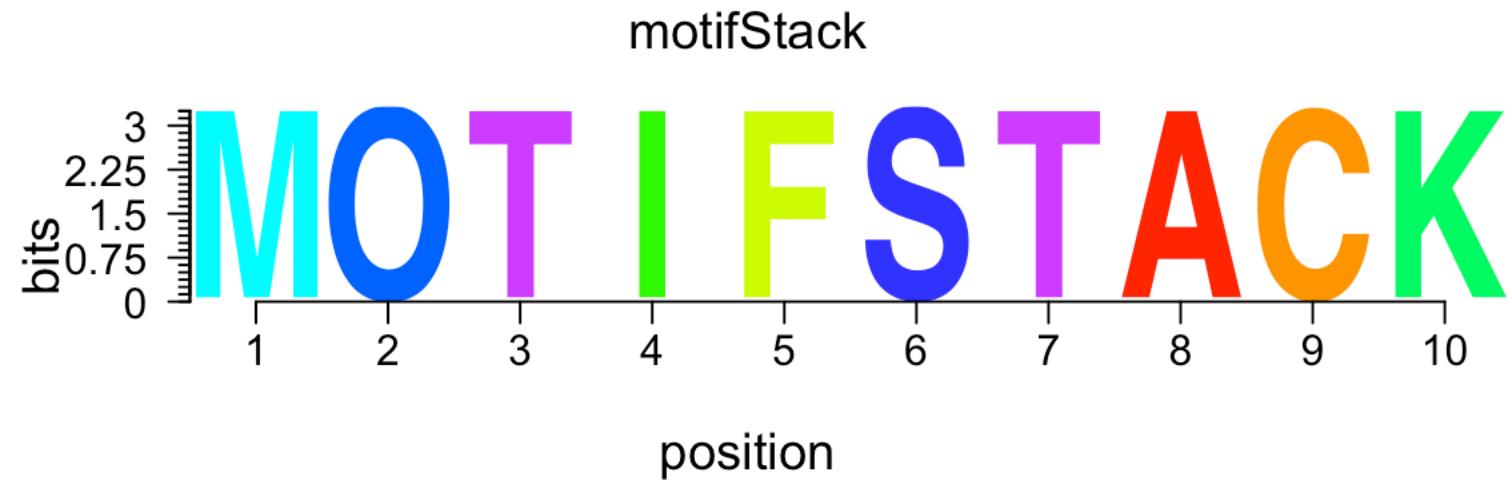
PLOT AN AMINO ACID (AA) SEQUENCE LOGO

```
library(Biostrings)
protein<-read.table(system.file("extdata", "motifStack", "irg.txt",
                                 package = "workshop2020"))
protein<-t(protein[,2:21])
rownames(protein) <- sort(AA_STANDARD)
protein_motif<-new("pcm", mat=protein, name="IRG",
                    color=colorset(alphabet="AA",colorScheme="chemistry"),
                    alphabet = "AA",
                    markers=list(new("marker", type="rect", start=c(1,27), stop=c(13,49),
                                   gp=gpar(col="#009E73", fill=NA, lty=2)))))
plot(protein_motif)
```



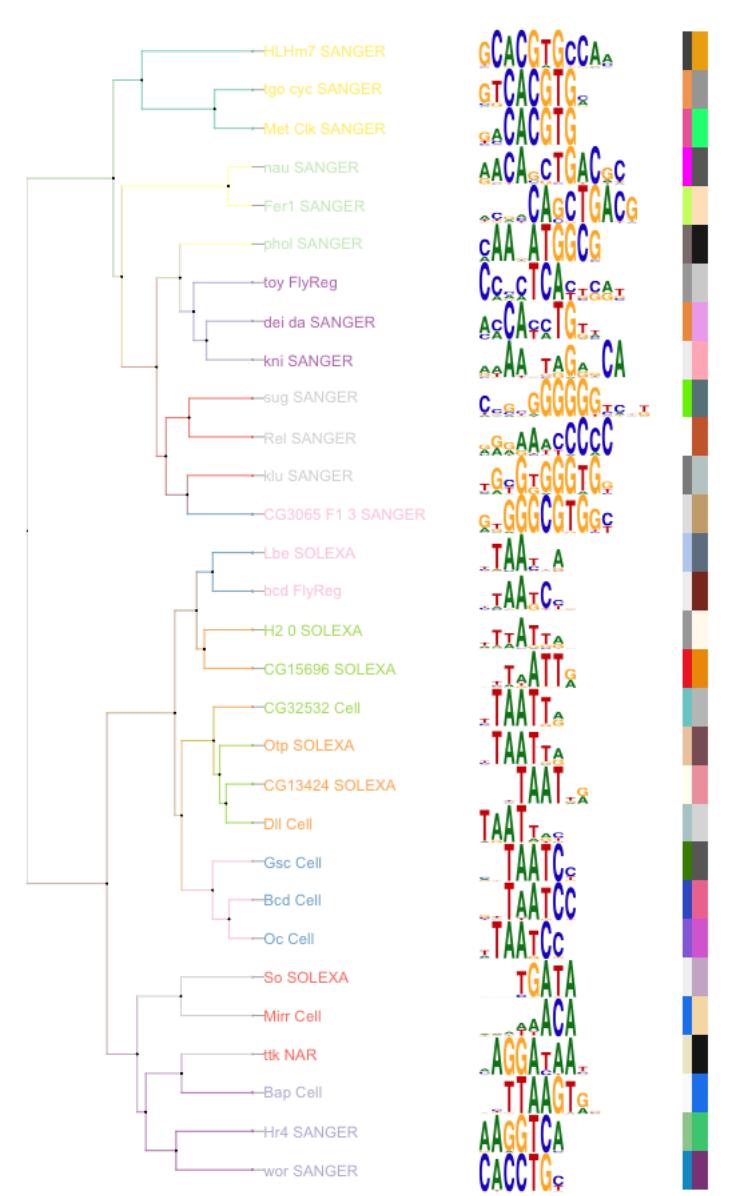
PLOT A CUSTOMIZED LOGO

```
m <- matrix(0, nrow = 10, ncol = 10,
            dimnames = list(strsplit("motifStack", "")[[1]],
                            strsplit("motifStack", "")[[1]]))
for(i in seq.int(10)) m[i, i] <- 1
ms <- new("pfm", mat=m, name="motifStack")
plot(ms)
```



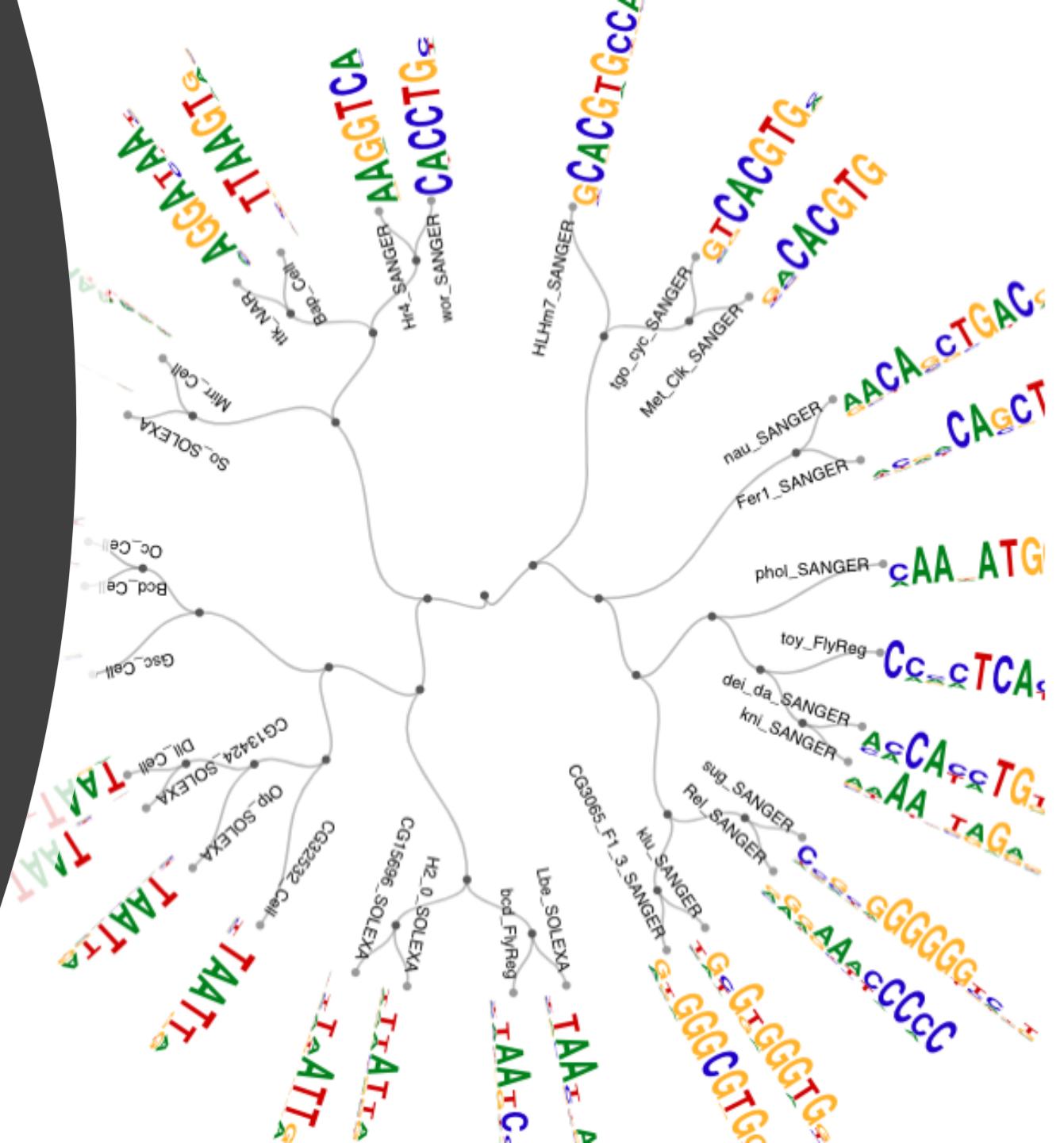
PLOT MULTIPLE SEQUENCE LOGOS

```
library(MotifDb); library(ade4); library(RColorBrewer)
matrix.fly <- MotifDb::query(MotifDb, "FlyFactorSurvey")
motifs2 <- as.list(matrix.fly)
## format the name
names(motifs2) <- gsub("(_[\\.\.0-9]+)*_FBgn\\\\d+$", "", elementMetadata(matrix.fly)$providerName)
names(motifs2) <- gsub("[^a-zA-Z0-9]", "_", names(motifs2))
motifs2 <- motifs2[unique(names(motifs2))]
## subsample motifs
set.seed(1); pfms <- sample(motifs2, 30)
## cluster the motifs
hc <- clusterMotifs(pfms)
## convert the hclust to phylog object
phylog <- ade4::hclust2phylog(hc)
## reorder the pfms by the order of hclust
leaves <- names(phylog$leaves)
pfms <- pfms[leaves]
## create a list of pfm objects
pfms <- mapply(pfms, names(pfms), FUN=function(.pfm, .name){
    new("pfm", mat=.pfm, name=.name)})
color <- brewer.pal(12, "Set3")
## plot the logo stack with pile style.
motifPiles(phylog=phylog, pfms=pfms,
            col.tree=rep(color, each=3), col.leaves=rep(rev(color), each=3),
            r.anno=c(0.02, 0.03), col.anno=list(sample(colors(), 30), sample(colors(), 30)),
            plotIndex=TRUE)
```



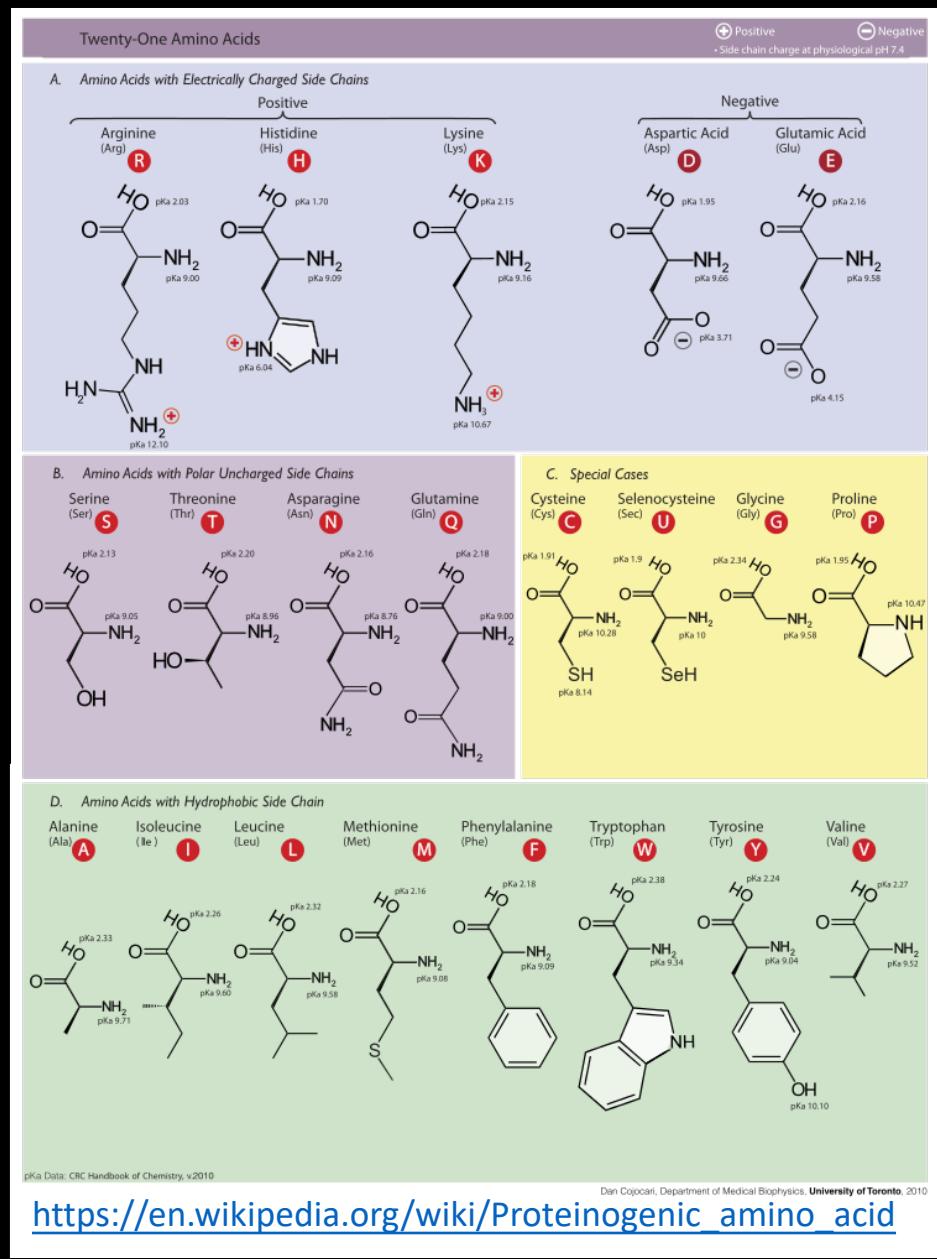
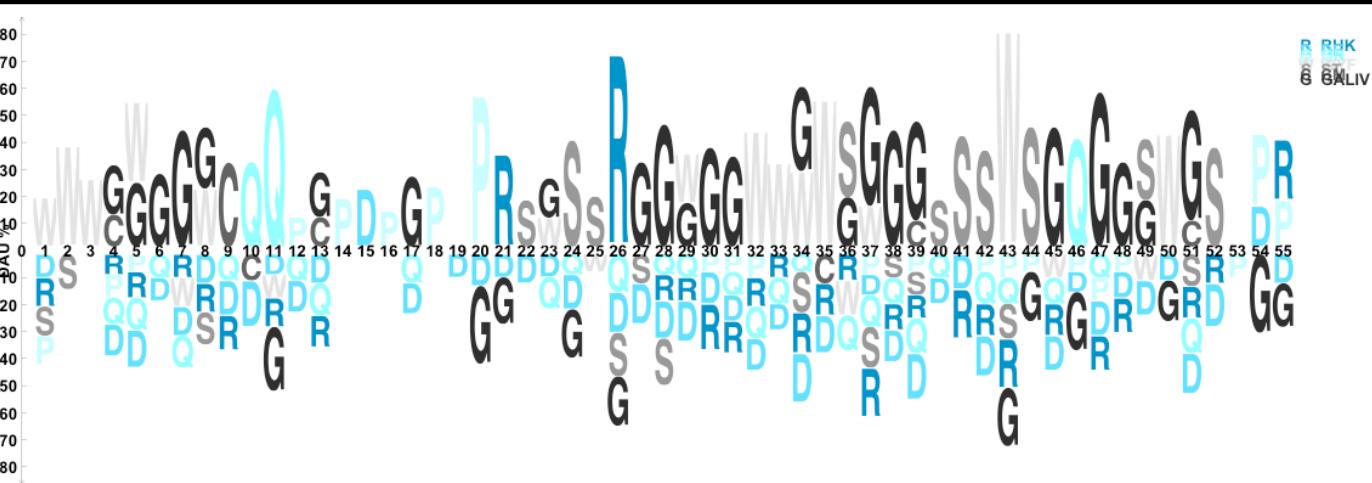
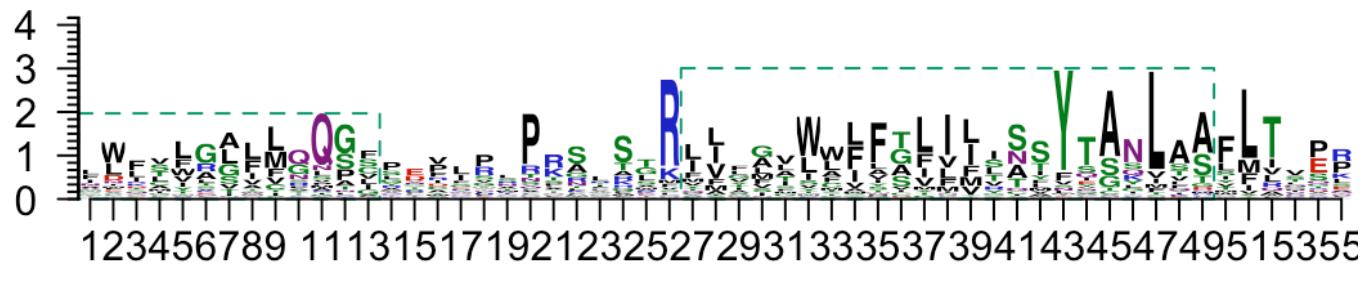
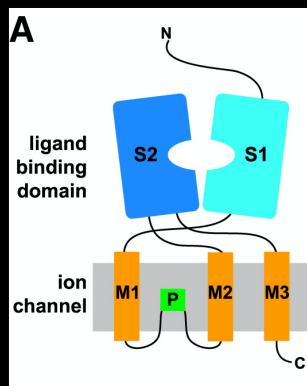
PLOT INTERACTIVE SEQUENCE LOGOS

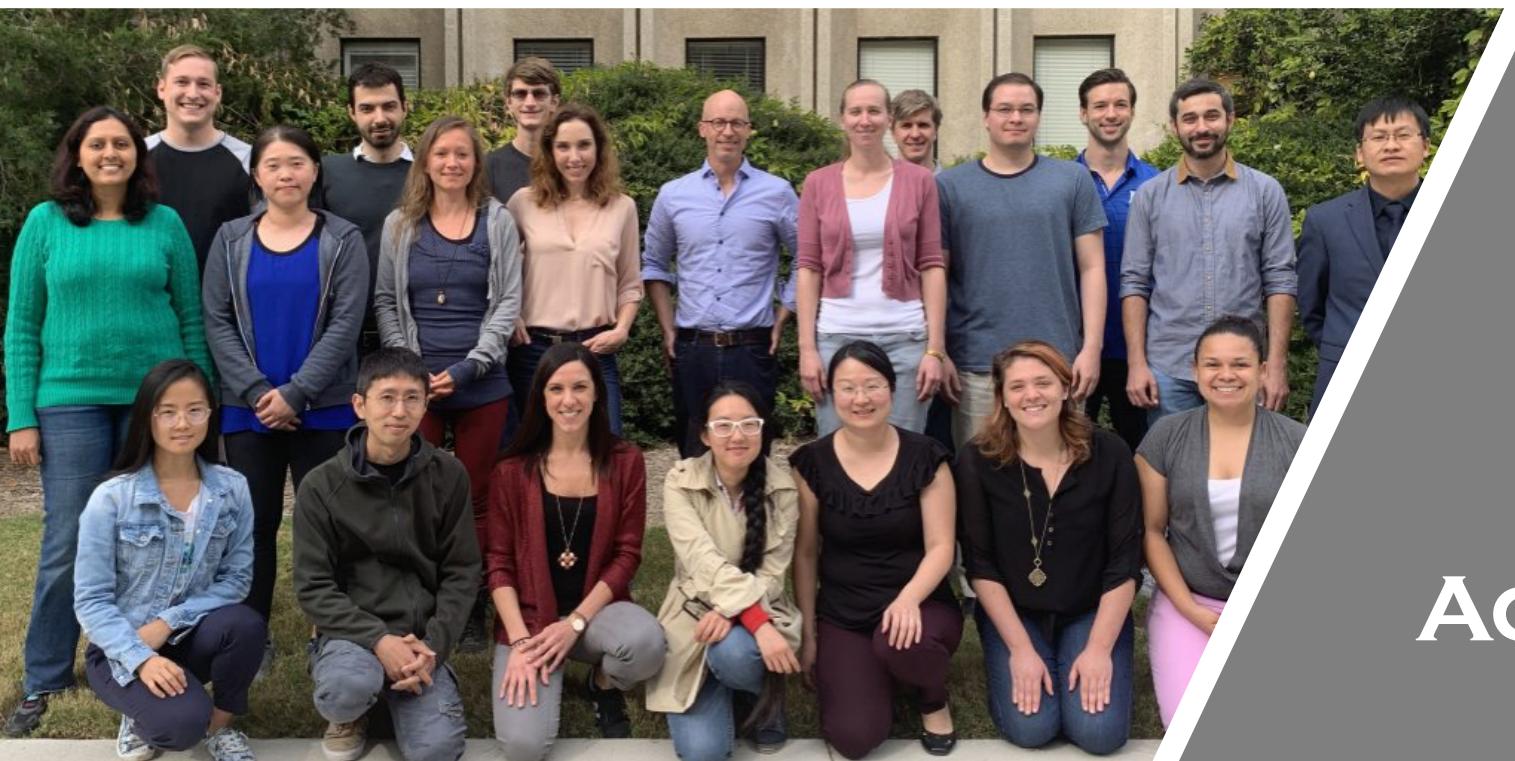
```
browseMotifs(pfms = pfms, phylog = phylog,  
           layout="radialPhylog",  
           yaxis = FALSE, xaxis = FALSE)
```



GOTO VIGNETTE

dagLogo





ACKNOWLEDGEMENT